



وزارة البحث العلمي والتعليم العالي
MINISTERE DE L'ENSEIGNEMENT SUPEREUR
ET DE LA RECHERCHE SCIENTIFIQUE
جامعة عبد الحميد بن باديس مستغانم
Université Abdelhamid Ibn Badis Mostaganem
كلية العلوم و التكنولوجيا
Faculté des Sciences et de la Technologie
DEPARTEMENT DE GENIE DES ELECTRIQUE



N° d'ordre : M2...../GE/2022

MEMOIRE DE FIN D'ETUDES DE MASTER EN ELECTRONIQUE

Option : Electronique des systèmes embarqués

Thème

Classification des lésions dans les tissus mammaires

Présenté par :

- HAMMOUDI Soltana
- KORCHI Malika

Soutenu le 04/07/ 2022 devant le jury composé de :

Président :	Mr. M.BENAOUALI	Université de Mostaganem
Examinatrice :	Mme. A.MEHIDI	Université de Mostaganem
Rapporteur :	Mme. K.BERRADJA	Université de Mostaganem

Année Universitaire 2021/2022

DÉDICACE

Tout d'abord, je tiens à remercier DIEU

De m'avoir donné la force et le courage de mener

À bien ce modeste travail.

Je tiens à dédier cet humble travail :

A mes très chers parents qui n'ont jamais cessé de nous encourager, pour leur soutien et tous les efforts qu'on me donné le long de mon parcours et je leur souhaite bonne santé et longue vie, que Dieu les protège

A ma sœurs **LIELA** mon frères **AMINE**

A tous les moments d'enfance passés avec vous mon frère ma sœur, en gage de ma profonde estime pour l'aide que vous m'avez apporté. Vous m'avez soutenu, réconforté et encouragé. Puissent nos liens fraternels se consolider et se pérenniser encore plus.

A TOUTE MA FAMILLE

A mes très chers amis

A tous ceux qui m'ont aidé de loin ou de près durant mes études.

MALIKA

DÉDICACE

Louange à Dieu tout puissant, qui m'a permis de voir ce jour tant attendu

Je dédie ce mémoire :

A mes très chers parents qu'ils trouvent en moi la source de leur fierté A qui je dois tout. Grâce à vous j'ai appris le sens du travail et de la responsabilité. Je voudrais vous remercier pour votre amour, votre générosité, votre compréhension... vos soutiens furent une lumière dans tout mon parcours. Aucune dédicace ne saurait exprimer l'amour l'estime et le respect que j'ai toujours eu pour vous. Ce modeste travail est le fruit de tous les sacrifices que vous avez déployés pour mon éducation et ma formation. J'implore le tout-puissant pour qu'il vous accordez une bonne santé et une vie longue et heureuse.

A ma sœur **IMENE** et mon frère **MOHAMED CHERIF**

A qui je souhaite un avenir radieux plein de réussite

A TOUTE MA FAMILLE

Aucun langage ne saurait exprimer mon respect et ma considération pour votre soutien et encouragements. Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur

A mes amies **Manel, Wissem, Asma, Ikram, Douae, Malika, Chaima** Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées, vous êtes pour moi des sœurs et des amies sur qui je peux compter. En témoignage de l'amitié qui nous unit et des souvenirs de tous les moments que nous avons passés ensemble, je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

SOLTANA

REMERCIEMENTS

En premier lieu, Nous remercions sans cesse Allah le tout miséricordieux d'être nos repères, de nous avoir guidés et éclairés la vie par le savoir, et de nous avoir donnés la volonté, la force et le courage afin de parvenir à élaborer ce modeste travail.

Nous exprimons toute notre gratitude à notre encadrante madame **K.BERRADJA**, pour nous avoir dirigées et guidées tout le long de ce travail. Nous la remercions pour sa haute compétence professionnelle, ses qualités humaines, ses conseils et remarques étaient très bénéfiques pour notre travail. Son soutien permanent ainsi que ses disponibilités pour l'achèvement de ce travail.

Nous remercions monsieur **M.BENAOUALI** Docteur à l'université de Mostaganem, pour avoir accepté de rapporter ce travail et de nous avoir fait l'honneur de présider le jury.

Nous tenons à remercier également madame **A.MEHIDI** Docteur à l'université de Mostaganem, pour avoir voulu partie de jury et accepté d'examiner ce travail.

Enfin, les mots les plus simples étant les plus forts, nous adressons toute notre affection à nos familles pour son soutien et ses encouragements permanents durant la préparation de cette thèse, et en particulier à nos parents pour leur compréhension, leur présence et leur patience durant cette période. Merci pour avoir fait de nous ce que nous sommes aujourd'hui.

Une pensée affectueuse va bien sûr à tous nos amis de l'université ou d'ailleurs, ainsi qu'à toute nos familles à tous ceux qui, de près ou de loin, ont contribué à l'accomplissement de ce modeste travail.

Résumé

Le cancer du sein est l'une des causes les plus fréquentes de décès chez les femmes. Une femme sur huit risque de développer la maladie au cours de sa vie. Il est essentiel de détecter la maladie tôt, avant qu'elle n'ait une chance de se propager. Du fait de son diagnostic tardif, il en résulte souvent par un traitement onéreux et coûteux. Grâce à la mammographie numérique et aux systèmes d'aide au diagnostic (CAD), les chances de survie sont fortement augmentées. Le défi consiste à détecter des cancers que les radiologues pourraient manquer. Cela en fait la principale direction de recherche pour notre projet. Dans ce travail, nous développons un système de détection et de classification des anomalies suspectes sur les mammographies. Le système qu'on propose se voit divisé en plusieurs étapes : une phase de pré traitement qui vise à réduire le bruit et améliorer le contraste. Ensuite, la segmentation dans laquelle nous introduisons le modèle déformable (actif contour) pour isoler les structures suspectes. Les tests expérimentaux ont montré l'efficacité de cet algorithme. Enfin, l'étape de classification, dans laquelle une classification SVM est effectuée, permettant d'identifier le type de l'anomalie observée (maligne ou bénigne).

Notre système a été testé sur des images procurées de la mini base de données MIAS

Mots-clefs : cancer du sein, segmentation par modèle déformable (actif contour), systèmes d'aide au diagnostic, classification SVM.

Table des matières

Introduction Générale	13
------------------------------------	----

Chapitre 01: Généralité sur le cancer du sein

1.1. Introduction	16
1.2. Anatomie du sein chez la femme	16
1.2.1. Composants anatomiques du sein.....	16
1.3. Cancer du sein	18
1.3.1. Types de cancers du sein et leurs caractéristiques.....	18
1.4. Dépistage du cancer du sein	19
1.4.1. La mammographie.....	19
1.5. Types d'anomalies dans le sein	21
1.5.1. Les calcifications	21
1.5.2. Les masses	22
1.6. Diagnostic du cancer du sein assisté par ordina-teur (CAOx).....	25
1.6.1. L'importance des systèmes de diagnostic par ordinateur.....	25
1.7. Les bases de données.....	26
1.7.1. Mammographie Image Analysis Society Digital Mammo-gram Database(MIAS).....	26
1.7.2. Digital Database For Screening Mammography(DDSM).....	27
1.7.3. Banco Web LAPIMO	27
1.8. Conclusion.....	28

Chapitre 02 : Traitement des mammographies

2.1. Introduction	30
2.2. Prétraitement des mammographies.....	31
2.2.1. Suppression des artefacts.....	32
2.2.2. Modifications d'histogrammes par une égalisation adaptative	32
2.2.3. Réduction du bruit	33
2.2.4. Amélioration de contraste.....	34
2.3. Segmentation des images mammographie	36

2.3.1. Segmentation par approche contour	37
2.3.1.1. Les Modèles déformables	38
2.3.1.1.1 Modèles déformables paramétriques	39
A. Contour actif (snake)	38
B. Méthode de résolution	38
C. L'algorithme du contour actif.....	42
2.4 Conclusion.....	44

Chapitre 03 : Les machines à vecteur support (SVM)

3.1. Introduction	46
3.2. Principe de fonctionnement	47
3.2.1. Apprentissage statistique	47
3.3. SVM principe de fonctionnement général en premier.....	48
3.3.1. Notions de base : Hyperplan, marge et support vecteur.....	48
3.3.2. SVMs binaires	49
3.3.2.1. SVM à marge dure.....	50
3.3.2.1. SVM à marge souple.....	54
3.3.2.1. Utilisation des noyaux.....	57
Exemples de noyaux	60
Exemple de séparation non linéaire.....	61
3.4. Conclusion.....	62

Chapitre 04 : Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

4.1. Problématique.....	64
4.2. Base de données.....	Erreur ! Signet non défini.
4.3. Conception du système d'aide au diagnostic assisté par ordinateur	65
4.3.1. Prétraitement des données	66
4.3.2. Segmentation	71
4.3.3. Descripteurs de forme.....	75
4.3.4. Classification des machines vectorielles SVM.....	75
4.4. Implémentation et présentation du CAOx.....	79
4.4.1. Environnement matériel et logiciel de programmation	79

4.4.2. Présentation de l'interface graphique de notre projet :.....	79
4.5. Conclusion.....	83
Cocclusion Générale.....	85
Annexe	
La morphologique mathématique	87
A.1. Introduction.....	87
➤ L'élément structurant	87
A.2. Les opérateurs morphologiques	88
➤ Erosion	88
➤ Dilatation	89
➤ Ouverture.....	89
➤ Fermeture	90
➤ Filtres Alternés Séquentiels (FAS).....	91
Bibliographie	93

Table de figure

Chapitre 01 : Généralité sur le cancer du sein

Figure 1.1 : Drainage lymphatique ganglionnaire des seins.....	17
Figure 1.2 : L'anatomie du sein chez la femme.....	18
Figure 1.3 : Exemple de mammographie.....	19
Figure 1.4 : Visualisation lever du bras pour oblique médio latérale (MLO)....	20
Figure 1.5 : Visualisation craniocaudal (CC).....	20
Figure 1.6: Les différentes formes possibles d'une masse.....	22
Figure 1.7: Les types de la densité possibles d'une masse.....	23
Figure 1.8: Les différents contours possibles d'une masse.....	24

Chapitre 02 : Traitement des mammographies

Figure 2.1 : Traitement d'image mammographie.....	30
Figure 2.2 : Exemple de Suppression des artefacts d'une image mammographie	32
Figure 2.3: Exemple d'image améliorée par égalisation d'histogramme adaptative.	33
Figure 2.4 : Élimination de bruit avec le filtre médian.....	34
Figure 2.5: Ouverture classique d'une image mammographie.....	35
Figure 2.6: Application du Filtre Alterné Séquentiel sur une image mammographie.	36
Figure 2.7 : Organigramme de la méthode de contour actif.....	43
Figure 2.8: Exemple de segmentation de masse par contours actifs (snakes)....	43

Chapitre 03 : Les machines à vecteur support (SVM)

Figure 3.1 : Exemple d'un hyperplan séparateur.....	49
Figure 3.2 : SVM binaire.....	49
Figure 3.3 : SVM binaire à marge souple.....	55

Figure 3.4 : Mal-adaptation de l'hyperplan aux problèmes réels de classification.....	58
Figure 3.5: Représentation idéale de la fonction de décision.....	58
Figure 3.6 : Transformation d'espace.....	59
Figure 3.7 : Exemple de changement de l'espace de données.....	61
 Chapitre 04 : Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (DAC)	
Figure 4.1 Les différentes étapes du traitement des mammographies.....	65
Figure 4.2 : Les différentes étapes du traitement des mammographies.....	66
Figure 4.3 : Suppression des artefacts : (a) et (b) images originelles, (c) et (d) après élimination des étiquettes.....	67
Figure 4.4 : Amélioration du contraste. (a) et (b) les images après la suppression des raies et des étiquettes, (c) et (d) amélioration de la qualité d'image(contraste) par égalisation adaptative d'histogramme.....	68
Figure 4.5 : Suppression du bruit : (a) et (b) images obtenues précédemment, (c) et (d) images filtrées.	69
Figure 4.6: Amélioration du contraste. (a) et (b) les images filtrées, (c) et (d) rehaussement de contraste par une ouverture.....	70
Figure 4.7 : Amélioration du contraste : (a) et (b) images filtrées par ouverture, (c) et (d) images avec rehaussement de contraste par filtre FAS.....	71
Figure 4.8 : (a)et (b) sont des images prétraitées, (c) et (d) sélectionnement de la zone d'intérêt par un marqueur semi – automatique.....	72
Figure 4.9: (a) et (b) des images avec un contour initial, (c) et (d) avec contour actif (snake).....	73
Figure 4.10 : Résultat de la segmentation par contour actif des deux images (a) mdb015 et (b) mdb202.....	74
Figure 4.11: menu de l'application.....	80
Figure 4.12 : résultat de l'élimination de bruit et amélioration du contraste de l'image mdb015.....	80
Figure 4.13 : résultat de l'élimination de bruit et amélioration du contraste de l'image mdb202.....	81
Figure 4.14: résultat de contour actif de l'image mdb015.....	81

Figure 4.15 : résultat de contour actif de l'image mdb202.....	82
Figure 4.16: résultat de la classification de l'image mdb015.....	82
Figure 4.17: résultat de la classification de l'image mdb202.....	83

Annexe

Figure A.1: Erosion d'une image mammographique.....	88
Figure A.2: Dilatation d'une image mammographie.....	89
Figure A.3: Ouverture d'une image mammographie par un élément structurant circulaire.....	90
Figure A.4 : Fermeture d'une image mammographique par un élément structurant.....	91
Figure A.5: Application du Filtre Alterné Séquentiel noir et blanc sur une image mammographique maligne.....	92

Table de Tableaux

Tableau 4.1: Calcul d'une matrice de confusion.....	76
Tableau 4.2: matrice de confusion de notre modèle SVM	77
Tableau 4.3: matrice de confusion de notre modèle SVM (2 ^{ème} cas).....	78

Introduction

Générale

Introduction générale

Le cancer du sein est le cancer le plus répandu et le plus mortel, menaçant la vie de la plupart des femmes dans le monde. Environ une femme sur huit est touchée par la maladie durant son existence. En Algérie, son incidence continue d'augmenter, avec près de 7500 cas enregistrés chaque année, dont environ 3500 décès, car les programmes de dépistage organisé n'existent toujours pas en Algérie. En raison de son diagnostic tardif, il en résulte souvent par un traitement onéreux et coûteux, avec un taux de mortalité élevé. Cependant, seul le fait de commencer le traitement dès le premier stade d'apparition de la tumeur peut réduire la mortalité et améliorer les chances de guérison de ce type de cancer. Pour assurer une détection précoce d'une telle tumeur, les radiologues ont dû augmenter la fréquence des mammographies, en particulier pour la tranche d'âge la plus touchée, ce qui, par exemple, en procédant à un dépistage systématique tous les deux ans, s'est révélé être une approche très efficace et réduire le taux de mortalité d'environ 35%. Deux ou quatre mammographies par patiente sont réalisées à raison d'une ou deux mammographies par sein, ce qui a entraîné une augmentation exponentielle du nombre de mammographies réalisées, rendant difficile la tâche d'interprétation des radiologues. Une seconde lecture peut améliorer le taux de détection du cancer du sein d'environ 15%. Avec l'augmentation du nombre de mammographies réalisées au cours des dernières décennies, divers efforts de recherche sont déployés pour détecter automatiquement les lésions (microcalcifications ou masses) ou pour fournir un deuxième avis sur la lésion détectée grâce à des systèmes de diagnostic assistés par ordinateur. Ceci a conduit au développement de plusieurs méthodes d'imagerie mammographique, toutes inspirées de la classification BI-RADS (Breast Imaging Reporting and Data System) de l'American College of Radiology (ACR) pour classer les masses en deux catégories, bénignes et malignes.

À cet effet, notre objectif est de développer un système d'aide automatique basé sur d'une part l'utilisation d'une méthode de segmentation et de détection automatique des masses à partir d'images mammographiques, et d'autre part l'utilisation des Support Vector Machines (SVM) pour classer à la fois la catégorie maligne ou bénigne pour que la patiente soit prise en charge dans les plus brefs délais.

Introduction générale

Ce mémoire s'articule autour de quatre chapitres qui présentent les différents aspects de ce travail

Le chapitre un est dédié à la définition des concepts de base de l'anatomie mammaire et à la description de ses principales composantes. En outre, ce chapitre comprend également une description générale des masses mammaires, de leurs caractéristiques et de leurs classifications.

Le chapitre deux est dédié à l'analyse des images de mammographie, d'une part les méthodes de prétraitement standard du traitement d'image sont décrites, telles que la réduction du bruit et l'amélioration du contraste, d'autre part la méthode des contours actifs pour la détection et l'extraction des masses.

Le chapitre trois est dédié à l'étude des machines à vecteurs de support (SVM) pour la classification.

Dans le chapitre quatre, un système de diagnostic assisté par ordinateur est proposé pour détecter les masses dans les images mammographiques et les classer comme bénignes ou malignes.

Enfin, une conclusion et les perspectives envisagées sont présentées en dernier lieu.

Chapitre 01



Généralités sur Le Cancer du Sein

Chapitre 1

Généralités sur le Cancer du Sein

1.1.Introduction

Le cancer du sein est le type de cancer le plus fréquent chez les femmes. Il affecte généralement les femmes ayant de plus de 50 ans. Maintenant, les femmes de tous âges peuvent avoir un cancer du sein. Ce dernier est aussi l'une des principales raisons de mortalité féminine.

Dans ce chapitre nous allons présenter les principaux éléments anatomiques du sein. Nous décrivons après le cancer ainsi que les types d'anomalies dans le sein. Ensuite, nous exposerons la procédure du dépistage du cancer du sein, puis nous continuerons avec les principes généraux d'acquisition en imagerie du sein. Cette introduction permettra de comprendre l'articulation des diverses étapes qui composent le système de diagnostic que nous proposons.

1.2.Anatomie du sein chez la femme

1.2.1. Composants anatomiques du sein

Le sein est une glande exocrine, sécrétant le lait, qui renferme la glande mammaire et qui se développe au cours de la vie de la femme.

- L'élément actif de la glande mammaire est constitué des alvéoles mammaires (ou acinus) qui sont regroupées en grappes pour former un lobule.
- La paroi des alvéoles est composée de cellules épithéliales sécrétoires et de cellules myoépithéliales contractiles.

Chapitre 1 Généralités sur le Cancer du Sein

- La glande mammaire contient environ une vingtaine de lobes. Chaque lobe regroupe 20 à 40 lobules dont les canaux se déversent vers un canal central : le canal lactifère [01].
- Les lobules sont essentiellement situés à la périphérie de l'organe (par rapport au mamelon), ce sont des groupes de glandes qui produisent le lait qui sera transporté au mamelon à partir des canaux galactophores [02].
- Le mamelon est la région située au centre de l'aréole et d'où sort le lait à une extrémité [03].
- L'aréole est la surface ronde, rosée ou brunâtre qui entoure le mamelon. Elle contient de petites glandes qui libèrent, ou sécrètent, une substance huileuse qui agit comme lubrifiant pour le mamelon et l'aréole [04].
- Le sein est parcouru de vaisseaux sanguins et de vaisseaux lymphatiques. Les ganglions lymphatiques de drainage du sein sont axillaires, sus-claviculaires, sous-claviculaires et mammaires internes (à l'intérieur du thorax). Ces derniers composent le système lymphatique qui aide à combattre les infections [05], comme montre la figure 1.1.

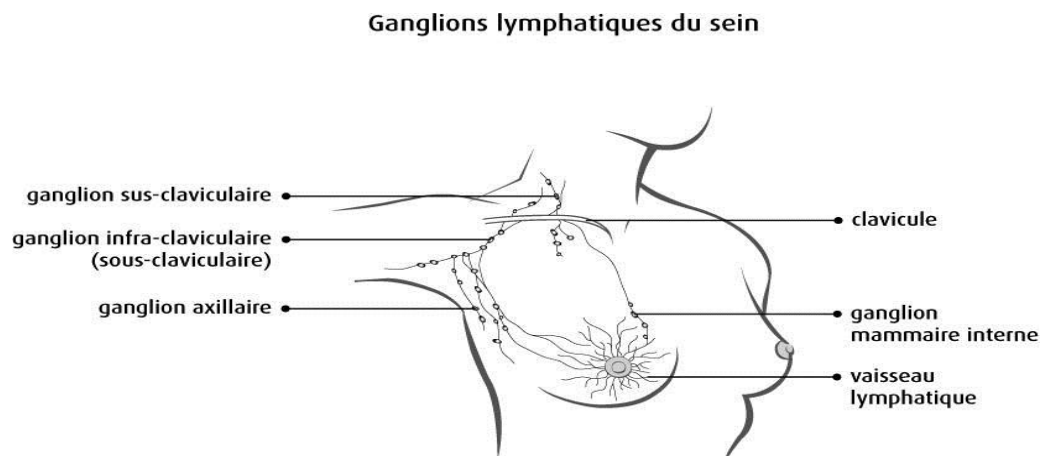


Figure 1.1 : Drainage lymphatique ganglionnaire des seins [06].

Enfin, les tissus conjonctifs entourent l'ensemble de la glande mammaire et les tissus graisseux lui donne forme et volume, et le tout sera recouvert par la peau.

La figure 1.2 détaille davantage ces différents composants mammaires.

Chapitre 1 Généralités sur le Cancer du Sein

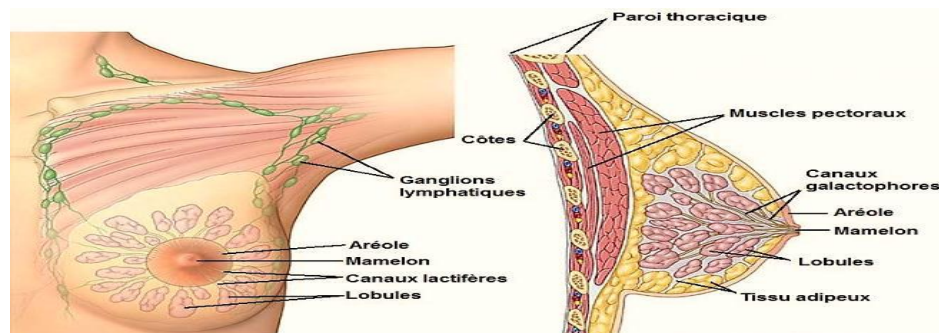


Figure 1.2 : L'anatomie du sein chez la femme [07].

1.3. Cancer du sein

Le cancer est une pathologie au cours de laquelle les cellules se reproduisent indéfiniment et anarchiquement en dehors des lois d'équilibre des tissus [08].

Les cancers du sein les plus fréquents (95 %) sont des adénocarcinomes, c'est-à-dire qui se développent à partir des cellules épithéliales (carcinome) de la glande mammaire (adénome) [09].

1.3.1. Types de cancers du sein et leurs caractéristiques

On distingue deux catégories de cancers du sein :

a) Tumeurs bénignes

Une tumeur bénigne du sein est une masse non cancéreuse, c'est-à-dire dénuée de potentiel métastatique et, de fait, destinée à rester localisée. On oppose les tumeurs bénignes aux tumeurs malignes du sein qui sont amenées à former des métastases se propageant aux autres organes du corps.

b) Tumeurs malignes

Les tumeurs malignes sont cancéreuses et agressives car elles envahissent et endommagent les tissus environnants. Lorsqu'une tumeur est soupçonnée d'être maligne, le médecin procédera à une biopsie afin de déterminer la gravité ou l'agressivité de la tumeur.

Chapitre 1 Généralités sur le Cancer du Sein

1.4. Dépistage du cancer du sein

1.4.1. La mammographie

La mammographie a pour but de déceler au plus tôt des anomalies avant même qu'elles n'aient provoqué des symptômes cliniques. Elle peut permettre, ainsi, de détecter des cancers bien avant qu'ils ne soient palpables.

L'examen consiste à comprimer à tour de rôle les deux seins puis à les exposer à une faible dose de rayons X, en utilisant le mammographe[10].

La mammographie ne permet pas toujours de donner d'emblée un diagnostic définitif : elle permet de voir s'il existe une anomalie dans le sein, mais elle ne permet pas de déterminer avec certitude s'il s'agit ou non d'un cancer.

La douleur lors de l'examen liée à la compression du sein entre deux plaques est fréquente mais de faible intensité. Le premier risque consécutif à la mammographie de dépistage est le sur-diagnostic. De 20 à 49 % des femmes participantes à un dépistage mammographie régulier auront au moins un résultat faussement positif après 10 examens.

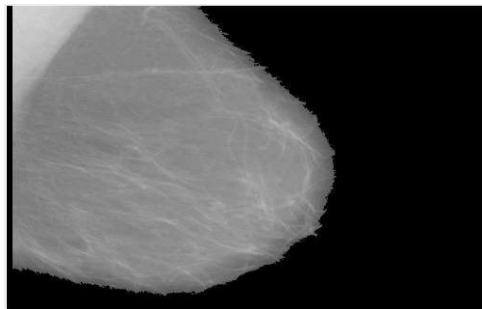


Figure 1.3 : Exemple de mammographie.

La mammographie numérique est une image électronique produite avant d'être sauvegardé sur ordinateur où elle subira une manipulation du degré de contraste dans les zones les plus dense notamment avant d'être visualisée. On distingue 2 positions de visualisation en mammographie numérique :

Chapitre 1 Généralités sur le Cancer du Sein

- **Lever du bras pour oblique médio latérale (MLO) :**

Lors de la visualisation du MLO, le patient est également amené à se pencher vers l'équipement pour une visualisation optimale des tissus. Le bras est tourné à 45 degré afin de démontrer la quantité maximale de tissu mammaire et de muscle pectoral. Parfois, l'angle est individualisé selon la taille de la poitrine avec une différence de 10 degré. L'autre sein du patient, non-imageur, est doucement pressé contre le corps et maintenu à l'écart [11] (Figure 1.4).

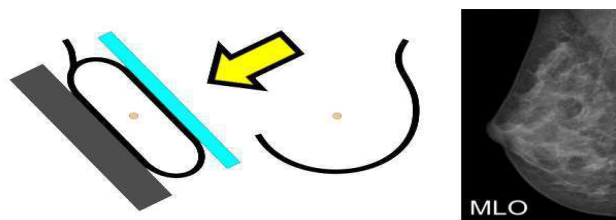


Figure 1.4 : Visualisation lever du bras pour oblique médio latérale (MLO) [12]

- **Craniocaudal (CC) :**

Pendant la vision CC, le patient est positionné de sorte que le mamelon se trouve approximativement au centre du détecteur. La patiente est obligée de se pencher vers l'appareil pour rapprocher le sein du détecteur. Le bas de la poitrine doit être soutenu et tiré vers le haut afin que les tissus les plus profonds et les plus bas soient inclus dans la vue, du côté en cours d'imagerie, est poussée vers le bas pour relâcher le muscle pectoral de manière à inclure le tissu mammaire dans le quadrant externe. La visualisation du muscle pectoral sur la vue en CC implique qu'aucun tissu le long de la paroi thoracique n'a été exclu (Figure 1.5).

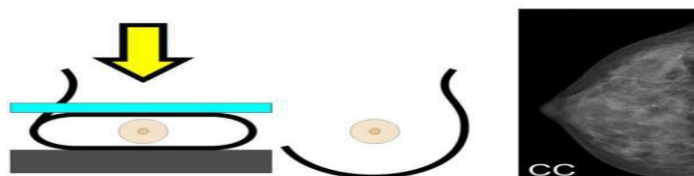


Figure 1.5 : Visualisation craniocaudal(CC) [13].

Chapitre 1 Généralités sur le Cancer du Sein

1.5. Types d'anomalies dans le sein

La généralisation du dépistage du cancer du sein, amène à découvrir beaucoup d'anomalies purement radiologiques. Parmi ces anomalies, on trouve :

- ✓ **Les calcifications.**
- ✓ **Les masses.**

1.5.1. Les calcifications :

Il existe deux types de calcifications :

- Les macro-calcifications.
- Les micro-calcifications.

Ils se traduisent, dans les clichés mammographies, par de petits points clairs.

✓ **Macro-calcification :**

Ce sont des gros dépôts de calcium dans le sein. Elles sont plus fréquentes chez les femmes âgées de plus de 50 ans. On les associe souvent à des affections bénignes ou des changements bénins qui se produisent dans le sein comme par exemple un vieillissement des artères du sein, des anciennes lésions, une infection des tissus du sein, des masses telles qu'un fibroadénome ou des kystes [14].

✓ **Micro-calcification :**

Ce sont des dépôts calcium minuscules ($< 0,5$ mm, généralement comprise entre 200 et 500 μ) qui ressemblent à des grains de sable dans le sein. Elles sont constituées de 10% des cristaux d'oxalate de calcium qui sont toujours bénignes et, 90% du phosphate de calcium qu'elles peuvent être bénignes ou malignes. Elles ne sont pas difficiles à voir en mammographie à cause de leur forte densité.

Quand elles sont nombreuses, de forme irrégulière et groupées, cela peut être un signe de cancer du sein [15].

Au cours du dépistage, l'analyse de la morphologie des micro-calcifications est très importantes afin d'en distinguer leurs bénignité ou leurs malignité, selon l'analyse de la morphologie de (Lanyi, 1985; D'Orsi et al. 2003) [16,17].

Chapitre 1 Généralités sur le Cancer du Sein

Les micro-calcifications arrondies ou ovales, uniformes dans leur taille et leur forme, sont souvent bénignes. Par contre, celles qui sont irrégulières et hétérogènes sont souvent malignes.

1.5.2. Les masses :

Une masse est une opacité importante occupant un espace dans le sein et vue comme une tâche blanche sur l'image mammographie. Il peut s'agir d'un kyste (collection liquidienne non cancéreuse) ou d'une lésion solide, qui peut correspondre à un cancer de sein.

Différents attributs permettent aux médecins de la décrire en vue de déterminer leur nature : sa forme, son contour et sa densité.

✓ La forme :

Selon la description du BIRADS (D'Orsi et al., 2003) [18], les masses mammaires peuvent avoir les formes suivantes :

1. Ronde : Il s'agit de masse sphérique, circulaire ou globuleuse.
2. Ovale : Elle présente une forme elliptique (ou en forme d'œuf).
3. Lobulée : La forme de la masse présente une légère ondulation.
4. Irrégulière : Cette appellation est réservée aux masses dont la forme est aléatoire.

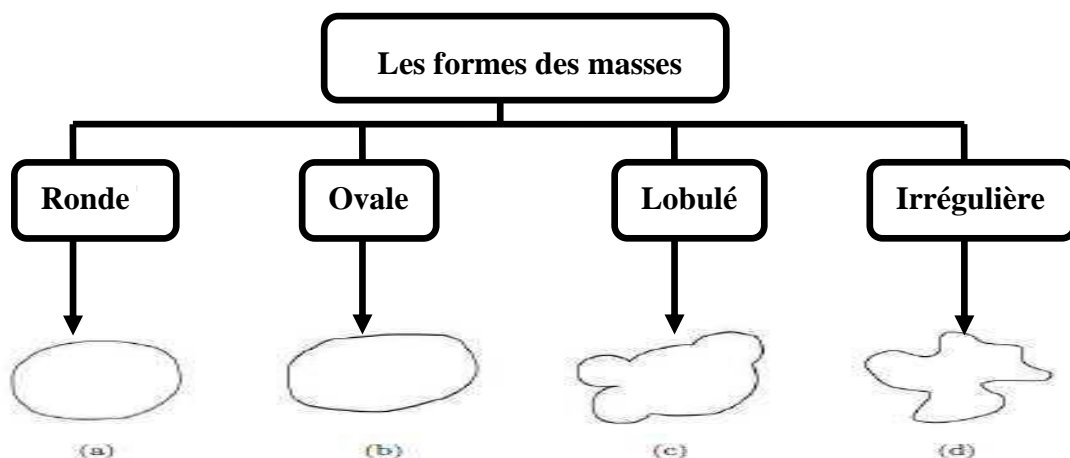


Figure 1.6: Les différentes formes possibles d'une masse [19].

Chapitre 1 Généralités sur le Cancer du Sein

✓ 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. Les tissus mammaires (graisseux, conjonctifs et glandulaires) changent au fur et à mesure en vieillissant. Selon (Wolf, 1976) [20], la densité du tissu mammaire est proportionnelle au risque de développement d'un cancer.

Afin de standardiser les comptes rendus mammographiques, la classification BIRADS (Breast Imaging Reporting And Data System) de l'ACR a défini 4 classes de la composition du sein [21].

➤ Les quatre types de la densité mammaire sont :

- Type 1 : Le sein est presque entièrement gras : sein clair de densité < 25%.
- Type 2 : Il y a des opacités fibro-glandulaires de densité entre 25 et 50%.
- Type 3 : Le tissu mammaire est dense et hétérogène de densité entre 50 et 75%.
- Type 4 : Le tissu mammaire est extrêmement dense de densité > 75%.

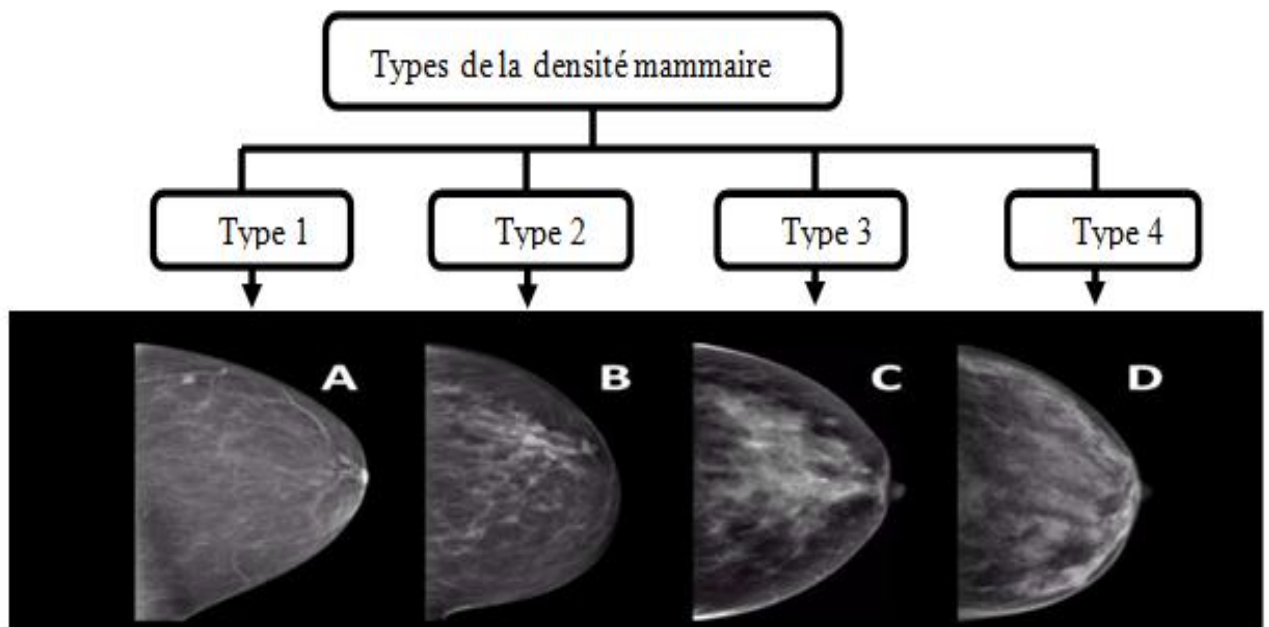


Figure 1.7: Les types de la densité possibles d'une masse [22].

Chapitre 1 Généralités sur le Cancer du Sein

✓ Le contour :

Les contours des masses mammaires sont les suivants :

- Circonscrit** : Il s'agit d'une transition brusque entre la lésion et le tissu environnant.
- 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.
- Micro-lobulé** : Dans ce cas, de courtes dentelures du contour créent de petites ondulations.
- Indistinct** : Dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre à une infiltration.
- Spiculé** : La masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.
- Distorsions architecturales** : ce sont des signes hautement suspects qui se traduisent par des structures linéaires ou des spicules qui convergent vers une même zone focale.

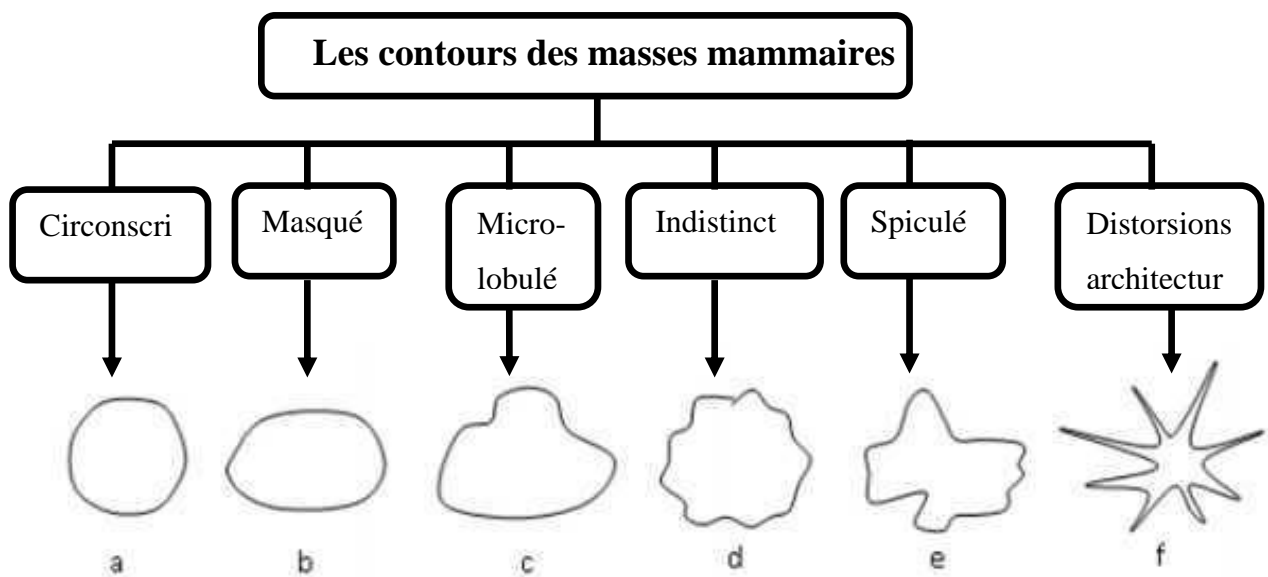


Figure 1.8: Les différents contours possibles d'une masse [23].

Chapitre 1 Généralités sur le Cancer du Sein

1.6. Diagnostic du cancer du sein assisté par ordina-teur (CAOx)

Le diagnostic des radiologues à propos des lésions mammaires est une tâche subjective qui peut donc être entachée d'erreur dont les conséquences peuvent s'avérer mortelles pour les patientes. Plusieurs raisons comme la fatigue, le manque d'expérience en imagerie du sein et la quantité d'images à analyser chaque jour rendent cette tâche fastidieuse et justifient les erreurs de diagnostic. A l'inverse, un système automatique de diagnostic assisté par ordinateur(CAOx) conçu sur un modèle qui prend en compte des descripteurs pertinents de la lésion mammaire fera toujours preuve de robustesse, ce qui pourrait aider à réduire le nombre de biopsies inutiles. En effet, il a été prouvé qu'après une biopsie, seulement moins du tiers des images mammographiques suspectes sont réellement des mammographies de lésions cancéreuses [Meyer et al., 1990]. Aussi, la conception des systèmes de CAOx pourrait aider les radiologues à diminuer le nombre de biopsies et éviter à la patiente une situation de stress inutile. [24] Un système automatique de CAOx employé dans un contexte de dépistage ou de diagnostic servira donc de second avis au radiologue et n'a absolument pas vocation à se substituer à ce dernier. De façon générale, le fonctionnement d'un système de CAOx est une suite de plusieurs étapes toutes aussi importantes les unes que les autres. Quand le système reçoit en entrée une image, il procède d'abord à un pré-traitement de celle-ci, c.-à-d élimination du bruit d'acquisition ou speckle et amélioration du contraste de l'image, afin d'obtenir la meilleure qualité d'image possible.

Cette étape est suivie de la détection des anomalies (p. ex. microcalcifications, opacité ou distorsions architecturales). Sur l'anomalie détectée sont ensuite extraits plusieurs attributs mathématiques pouvant aider à décrire au mieux la nature de la lésion. Pour finir, les descripteurs extraits sont ensuite classifiés en s'appuyant sur des algorithmes adaptés à cette tâche, afin de décider de la nature de la masse.

1.6.1. L'importance des systèmes de diagnostic par ordinateur

Même si la lésion est très suspecte en mammographie ou échographie, le diagnostic préopératoire est aujourd'hui indispensable pour deux raisons :

Chapitre 1 Généralités sur le Cancer du Sein

- Il peut éviter de nombreuses chirurgies inutiles s'il ne s'agit pas d'un cancer.
- Il permet d'optimiser d'emblée la chirurgie s'il s'agit d'un cancer du sein : si ce cancer est de petite taille par exemple, une intervention sur un ganglion (ex : ganglion sentinelle) pourra être programmée et permettra à la patiente d'éviter un curage axillaire. Dans d'autres cas par exemple, le traitement pourrait être d'abord médical (néo-adjuvant) pour permettre une chirurgie ensuite moins lourde.

Dans de très rares cas, ce diagnostic pré-opératoire n'est pas possible en raison de la localisation des images radiologiques qui sont, par exemple, non accessibles à une biopsie. La chirurgie permettra de poser le diagnostic [25].

1.7. Les bases de données

1.7.1. Mammographie Image Analysis Society Digital Mammo-gram Database(MIAS)

Mammographic Image Analysis Society (MIAS) [Suckling et al., 1994] est la plus ancienne base d'images publiquement disponible ; elle a été conçue au Royaume-Uni en 1994 et est encore largement utilisée dans l'état-de-l'art. La MIAS contient 161 cas pour un total de 322 images numérisées obtenues en incidence MLO, sur lesquelles sont présentes toutes les pathologies à savoir les lésions bénignes et malignes, mais également les images normales. Cette base dispose d'un nombre important de masses spiculées, ainsi que d'une information relative à la densité du sein ; toutefois, la classification des masses réalisées par les radiologues ayant évalué cette base ne respecte pas les standards de l'ACR. Cependant, avec l'augmentation de l'usage de la classification BI-RADS de l'ACR, certains auteurs se sont essayés à une classification des masses de la MIAS afin qu'elle puisse correspondre aux standards en vigueur en imagerie du sein. Un autre défaut de la MIAS est l'annotation de ces images, qui consiste à indiquer le centre et le rayon de la région d'intérêt, c'est-à-dire la région où se situe la pathologie.

Chapitre 1 Généralités sur le Cancer du Sein

Ce genre d'annotation est considérée comme insuffisant pour certaines études comme la segmentation des masses, où l'on souhaite que toutes les lésions circonscrites ou spiculées soient manuellement segmentées afin de faire une comparaison très précise avec les méthodes automatiques. Pour finir, l'autre inconvénient de la MIAS, c'est la résolution à laquelle les images sont numérisées et qui fait que cette base ne convient pas pour des expériences liées à la détection des micro-calcifications [26].

1.7.2. Digital Database For Screening Mammography(DDSM)

Constituée en 1999 aux Etats-Unis (USA), la Digital Database for Screening Mammography (DDSM) est la plus grande base publique de mammographies. Elle contient 2620 cas incluant deux images (incidences MLO et CC) de chaque sein, pour un total de 10480 images où se retrouvent toutes les pathologies, des images normales à celles contenant des masses bénignes et malignes. Pour chacune de ces images, les informations relatives à l'âge de la patiente, la densité du sein et une classification respectant le BI-RADS sont fournies.

A la différence de la MIAS, les annotations concernant les pathologies sont faites manuellement par les radiologues, qui encerclent de façon un peu grossière la région suspecte. Toutefois, des travaux de recherche ont souligné que la DDSM n'était pas adéquate pour la validation des algorithmes de segmentation, en raison du manque de précision de la segmentation manuelle [27].

1.7.3. Banco Web LAPIMO

Plus récente que les deux précédentes bases, la Banco Web LAPIMO est une base créée en 2010 au Brésil. Elle dispose de 320 cas avec 1473 images (MLO et CC) contenant toutes les pathologies. Les antécédents de la patiente, de même qu'une classification BI-RADS sont fournis. En ce qui concerne les masses, seules certaines images sont annotées sous forme de ROI ; mais toutes les pathologies disposent d'une description textuelle [28].

Chapitre 1 Généralités sur le Cancer du Sein

1.8. Conclusion

Dans ce chapitre, nous définissons quelques informations générales sur les glandes mammaires et leurs systèmes d'acquisition. Une présentation concise de l'anatomie du sein, ainsi qu'une étude des caractéristiques pathologiques du sein telles que les microcalcifications, les macrocalcifications et les masses y sont présentées également.

Dans le chapitre suivant, nous présenterons les différentes étapes du traitement des images mammographiques afin d'améliorer la qualité des mammographies pour des performances de segmentation et de classification optimales.

Chapitre 02



Traitement des Images de Mammographies

Chapitre 02

Traitement des Images de mammographies

2.1. Introduction

Les mammographies sont des images qui n'offrent pas un bon contraste entre les tissus mammaires sains et malins ; cela peut s'expliquer par le fait qu'il y a peu de variation d'atténuation des rayons X entre ces tissus, en particulier chez les jeunes femmes présentant un parenchyme mammaire moyen très dense. Le traitement de l'image se résume comme suit :

- Le prétraitement comprend diverses opérations visant à améliorer la qualité de l'image et à faciliter sa segmentation. Ces opérations sont principalement l'amélioration du contraste, la modification de l'histogramme et la réduction du bruit.

- La segmentation est un processus de bas niveau qui consiste à diviser une image en sous-ensembles appelés régions. La segmentation est alors obtenue en extrayant les composantes connexes de pixels appartenant à une même classe. Par conséquent, une région sera constituée de pixels de la même classe.

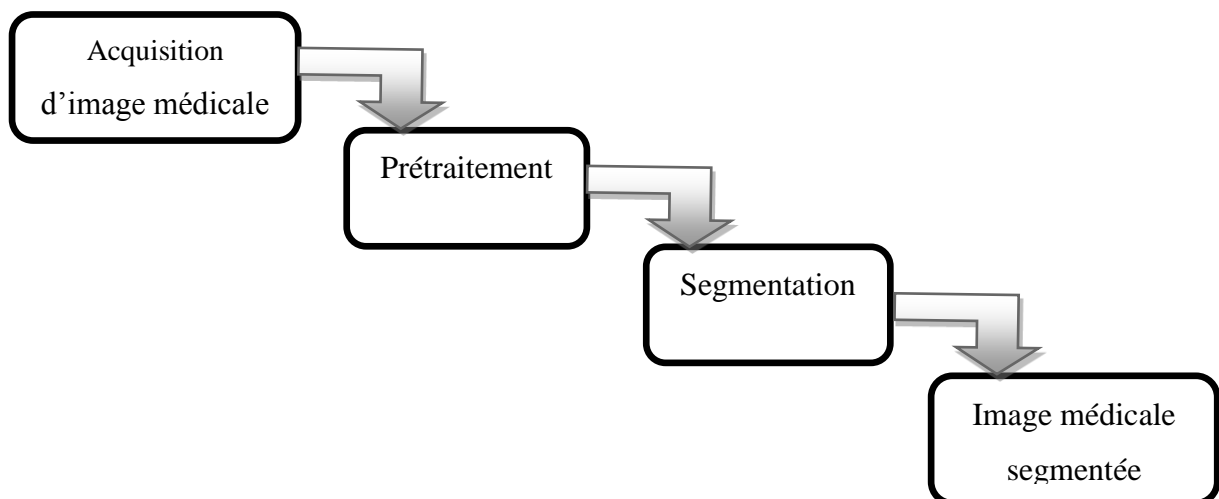


Figure 2.1 : Traitement d'image mammographie [29]

Chapitre 02 Traitement des Images de mammographies

2.2. Prétraitement des mammographies

Les techniques standard de réduction du bruit et d'amélioration de contraste ont d'abord été appliquées aux mammographies, puis, on s'est rendu compte que, hormis l'amélioration apportée à l'image, ces approches contribuaient à amplifier de façon significative le bruit présent dans l'image. Ensuite, certaines méthodes plus spécifiques sont suggérées qui aident souvent à améliorer la qualité des mammographies, telles que l'explication qui indique également que l'utilisation de ces techniques dans un cadre clinique peut réduire de moitié la dose de rayonnement lors de l'enregistrement de la mammographie. Cependant, le vrai problème des algorithmes de prétraitement est que certaines zones peuvent être très bien traitées tandis que d'autres peuvent être plus ou moins traitées.

Sans traitement initial, de nombreux détails de la lésion peuvent être manqués ; dans ce cas, certaines tumeurs peuvent ne pas être détectées, conduisant à des résultats faussement négatifs. D'autre part, une amélioration excessive du contraste peut ajouter des détails ou des artefacts à l'image, donnant au radiologue l'impression qu'une tumeur est présente, entraînant des biopsies inutiles et une augmentation des fausses images. A cela s'ajoute le fait que les approches locales et multi-échelles nécessitent la spécification de plusieurs paramètres (par exemple taille, et type de filtres, seuils et différentes résolutions) qui ont un impact important sur la qualité des résultats obtenus [30]. Les techniques d'amélioration des images offrent de nombreuses possibilités d'améliorer leur qualité visuelle.

Le choix approprié de ces techniques est fortement influencé par l'importance du détail de l'image et le type d'information qui doit être amélioré et/ou éliminé. Cette étape vise à améliorer la qualité de l'image traitée, en appliquant d'abord un filtre médian servant à réduire le bruit tout en préservant les contours de l'image, suivi d'un filtre morphologique pour éliminer les intensités dans l'image et donc un bon résultat. Le bon choix de ces techniques est fortement influencé par l'importance des détails de l'image et le type d'information que l'on souhaite améliorer et/ou éliminer.

Cette étape consiste à améliorer la qualité de l'image traitée, qui fonctionne d'abord en appliquant un filtre médian pour réduire le bruit tout en préservant les bords de l'image.

Chapitre 02 Traitement des Images de mammographies

2.2.1. Suppression des artefacts

La mammographie numérique contient généralement deux zones distinctes : la zone exposée du sein et la zone non exposée, qui est le fond de l'œil. Ce fond d'œil peut contenir des artefacts que le système visuel humain peut facilement ignorer lors de l'interprétation, par contre le système automatique. Vous devez d'abord identifier et classer les artefacts qui conduisent à des interprétations erronées en analysant les images utilisées (Figure 2.2). Les mammographies sont généralement marquées d'une sorte d'étiquette d'identification permanente avec, qui contient des informations sur le test effectué. Ces étiquettes sont des indicateurs radio-opaques qui montrent la latéralité mammographique (R/L, Right/Left) et des indicateurs de projection MLO/CC de l'axe X de l'image mammographique.

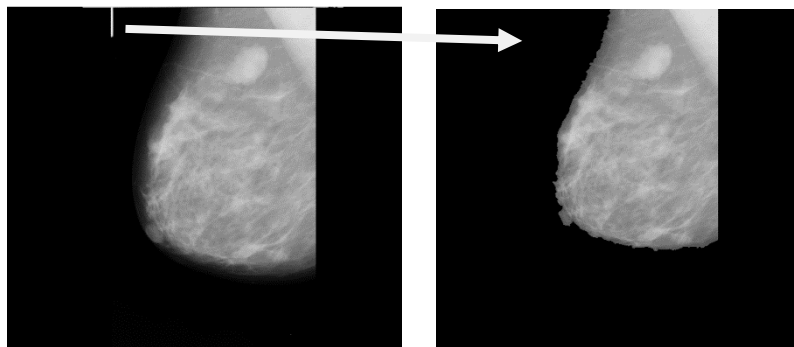


Figure 2.2 : Exemple de Suppression des artefacts d'une image mammographie.

2.2.2. Modifications d'histogrammes par une égalisation adaptative

Afin d'améliorer les caractéristiques de l'image (en général, pour accentuer les contrastes), une approche générale consiste à appliquer une fonction qui attribue une nouvelle valeur à chaque valeur d'intensité de l'image, cette fonction modifie l'histogramme de l'image.

Son principe consiste à appliquer sur chaque pixel ainsi que sa région contextuelle une égalisation d'histogramme. Cette région représente en effet les pixels voisins entourant le pixel traité [31]. La figure 2.3 montre l'image mdb015 améliorée par égalisation d'histogramme adaptative avec une implémentation sous Matlab.

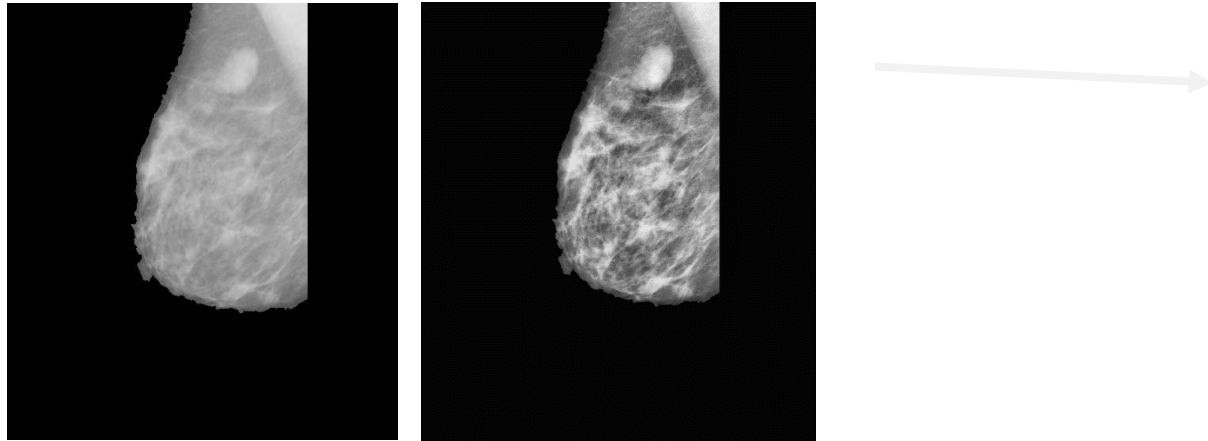


Figure 2.3: Exemple d'image améliorée par égalisation d'histogramme adaptative.

2.2.3. Réduction du bruit

La réduction du bruit est réalisée en appliquant des filtres médians pour éliminer le bruit causé par le dispositif d'acquisition (la mammographie dans notre cas)

- **Filtre médian**

Le filtre médian est un filtre numérique non linéaire, souvent utilisé pour la réduction de bruit en conservant les contours de l'image [32]. L'idée principale du filtre médian est de remplacer chaque entrée par la valeur médiane de son voisinage. Par exemple, si on considère ces neuf pixels, dont une valeur aberrante, ici (86)

5	6	7
6	86	8
7	8	9

Le filtre médian va considérer les valeurs du voisinage par valeurs croissantes :

5	6	6	7	7	8	8	9	86
---	---	---	---	---	---	---	---	----

Chapitre 02 Traitement des Images de mammographies

Et prendre la valeur médiane, ici la valeur 7. La sortie du filtre donnera donc :

5	6	7
6	7	8
7	8	9

Ce qui a permis de remplacer la valeur aberrante par une valeur "de consensus" entre les valeurs voisines.

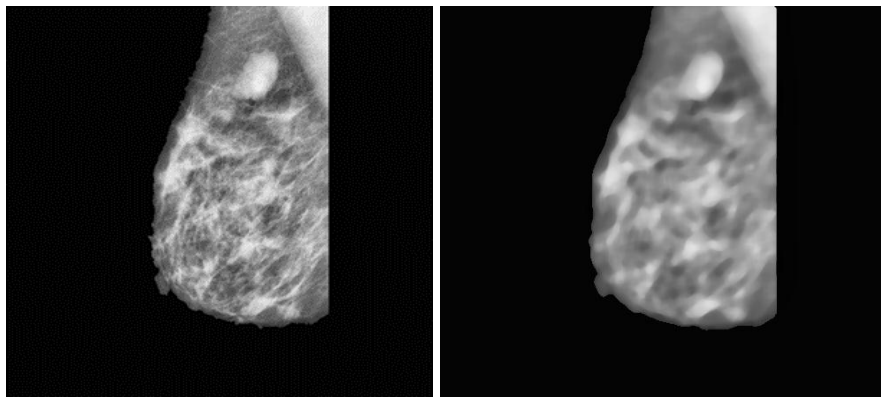


Figure 2.4 : Élimination de bruit avec le filtre médian.

2.2.4. Amélioration de contraste

Le Contraste dans une image peut être définie comme :

Écart-type des variations des niveaux de gris dans l'image. Dans le cadre des images médicales, ce traitement a pour but de mettre en avant les régions d'intérêt en atténuant les autres afin de faciliter le diagnostic. Nous présentons amélioration du contraste par :

1) L'ouverture morphologique classique :

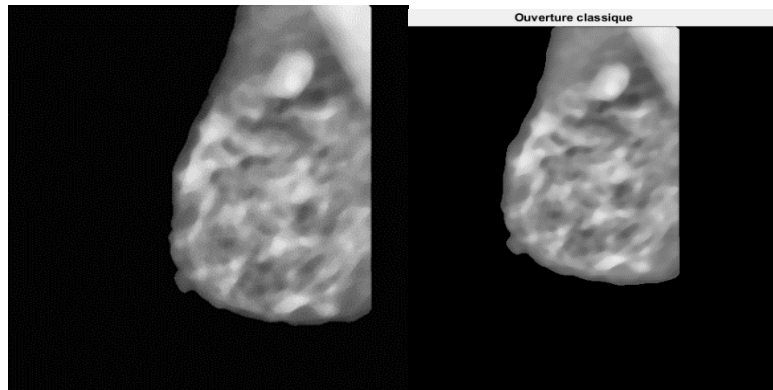
L'ouverture est un filtre morphologique qui est généralement utilisée pour restaurer ou récupérer l'image d'origine dans la mesure du possible.

L'ouverture morphologique est égale à la composition d'une érosion suivie d'une dilatation. C'est un opérateur croissant et idempotent c'est à dire le résultat ne

Chapitre 02 Traitement des Images de mammographies

change pas si on applique plusieurs fois l'opérateur, il suffit de l'appliquer une seule fois.

- Elle a pour effets : de faire disparaître les petites particules et de séparer les grosses particules aux endroits où elles sont plus fines.



a) Image mammographique filtrée b) Image mammographique avec ouverture

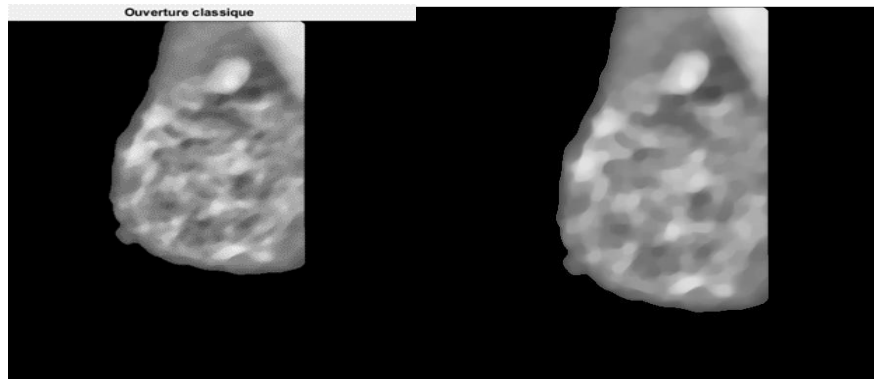
Figure 2.5: Ouverture classique d'une image mammographie.

2) Filtres Alternés Séquentiels (FAS)

A partir des filtres morphologiques élémentaires, tels que les ouvertures et les fermetures, on définit des filtres morphologiques plus évolués. Parmi ceux-ci, les Filtres Alternés Séquentiels (FAS) qui sont sans doute les plus populaires. Ils permettent de réduire la complexité d'une image sans dégrader les contrastes les plus importants.

Les ouvertures suppriment les détails blancs sans modifier les parties sombres de l'image tandis que les fermetures font l'inverse. Les filtres alternés séquentiels (FAS) adoptent un comportement plus symétrique par une composition alternée d'ouvertures et de fermetures.

Chapitre 02 Traitement des Images de mammographies



a) Image mammographique avec ouverture b) Image mammographique filtrée par un FAS

Figure 2.6: Application du Filtre Alterné Séquentiel sur une image mammographie.

2.3. Segmentation des images mammographie

La segmentation est un problème omniprésent dans le traitement d'image et la vision par ordinateur où le but est d'isoler les objets de l'arrière-plan de l'image. Dans le cadre de la conception du système CAOx, la segmentation joue un rôle central, tout comme les autres étapes, à savoir la description de la masse, la détection des structures d'intérêt telles que les spicules, et la classification des différentes lésions qui en dépendent fortement [33]. Par conséquent, plus la segmentation est précise, meilleure est la description des propriétés du contour, de la texture et de la forme de la masse, ce qui conduira à une bonne classification de la tumeur. La particularité de la segmentation est qu'aucune méthode ne fonctionne pour toutes les applications. A ce titre, il a fait l'objet de plusieurs publications dans lesquelles les chercheurs ont tenté d'apporter différentes solutions, souvent en tenant compte de la nature des images traitées.

De façon générale, les méthodes de segmentation sont regroupées en deux catégories que sont les approches supervisées et non-supervisées. Les approches non-supervisées réalisent une partition de l'image en plusieurs régions distinctes qui satisfont des propriétés liées au niveau de gris des pixels et à la texture. Ces approches sont également subdivisées en trois groupes que sont les méthodes basées régions, celles basées contour et celles basées clustering. Pour leur part, les approches supervisées c-à-d nécessitant une intervention humaine – sont fondées

Chapitre 02 Traitement des Images de mammographies

sur l'utilisation d'une information a priori sur la nature de l'objet à extraire et éventuellement le fond de l'image.

2.3.1. Segmentation par approche contour

Un contour est un ensemble de pixels formant une frontière entre deux ou plusieurs régions voisines.

L'épaisseur d'un contour est d'un ou plusieurs pixels et il est défini par une variation « brusque ». La détection des contours dans une image est un problème très étudié dans le domaine du traitement et de l'analyse de l'image. De façon générale, un contour est défini comme le lieu de variation significative de l'information "niveau de gris". Par conséquent, trouver les contours dans une image revient à évaluer la variation du niveau de gris en chaque pixel de l'image. Les méthodes d'extraction de contours s'appuient sur la détection des discontinuités dans l'image et peuvent être divisées en trois classes :

- Les méthodes dérivatives.
- Les méthodes analytiques.
- Les méthodes basées sur les contours actifs

2.3.1.1. Les Modèles déformables

Les modèles déformables connus aussi sous les noms «snacks» et «contours actifs» ont été introduits par Kass [Kass et al., 1987] et Terzopoulos [Terzopoulos and Fleischer,1988]. Initialisé proche du contour, le modèle déformable évolue itérativement de manière à s'ajuster au contour de l'objet à détecter. Il se présente sous la forme d'une courbe $C = [x(t),y(t)]$ paramétrée par son indice curviligne t . La forme et la position initiale de la courbe déformable sont fournies de manière manuelle ou automatique selon l'algorithme proposé. Le snake s'équilibre et s'immobilise lorsqu'il minimise son énergie. Le critère à optimiser est une somme pondérée d'énergies qui traduisent des forces de natures différentes. Principalement, on cite les forces internes qui contrôlent la régularité de la courbe recherchée et les forces externes qui sont exercées par l'image sur le contour actif.

Chapitre 02 Traitement des Images de mammographies

2.3.1.1.1 Modèles déformables paramétriques

A. Contour actif (snake) :

La méthode de contours actifs est une méthode semi-interactive qui consiste à positionner dans l'image, plus précisément au voisinage de la forme à détecter, un contour initial qui subira une déformation sous l'effet de plusieurs forces telle que :

- Une énergie interne $E_{interne}$ permettant de régulariser le contour
- Une énergie potentielle E_{image} reliée à l'image
- Une énergie externe $E_{externe}$ reliée aux contraintes particulières que l'on peut ajouter.

Ces énergies vont permettre au contour actif d'évoluer pour prospecter la position d'énergie minimale qui sera ainsi un arrangement entre les diverses contraintes du problème.

B. Méthode de résolution

B.1. Equations de bases

L'équation qui caractérise une courbe paramétrée dans une image est l'équation de Snake (2.1) :

$$v(s) = [x(s), y(s)]^t, s \in [a, b] \quad (2.1)$$

Où S est généralement l'abscisse curviligne (longueur de la courbe).

Le contour initial est défini par v_0 .

L'opérateur cherche à définir la vitesse d'évolution de la courbe, cette dernière évolue vers un minimum local correspondant aux contours des objets.

▪ L'expression de l'énergie totale

L'énergie totale du Snake correspond à l'intégrale de l'ensemble des énergies exercé sur un domaine bien définie et qui est donnée par la relation (2.2) :

$$E_{snake} = \int_a^b [E_{interne}(v(s)) + E_{image}(v(s)) + E_{externe}(v(s))] ds \quad (2.2)$$

Chapitre 02 Traitement des Images de mammographies

▪ Energie interne

L'énergie interne égale à la première et à la seconde dérivée de la courbe paramétrée représentant le Snake :

$$E_{interne} = \alpha(s)E_{elastique} + \beta(s)E_{courbure} = \alpha(s)\left(\frac{dv}{ds}\right)^2 + \beta(s)\left(\frac{d^2v}{ds^2}\right)^2 \quad (2.3)$$

Où

α : Facteur d'élasticité

β : Facteur de rigidité du contour permettant ainsi d'obtenir des courbes plus ou moins lisses.

▪ L'expression de l'énergie potentielle

L'énergie potentielle liée à l'image représente les éléments sur l'image vers lesquels on veut attirer le Snake. Cette énergie par la formule suivant :

$$E_{image} = -\lambda(s)|\nabla I(v(s))|^2 \quad (2.4)$$

Avec :

λ est facteur qui dépend de l'image I initiale

∇ est l'opérateur gradient.

On peut faire précéder le gradient d'un filtrage passe-bas de l'image permettant d'obtenir des contours moins bruités et d'augmenter leur zone d'influence.

▪ L'expression de l'énergie externe

L'énergie externe (contraintes) est définie par l'opérateur selon les caractéristiques du problème étudié. Dans le cas d'absence de contraintes extérieures, cette énergie pourra être négligeable, l'énergie totale se résume à la relation (2.5) :

$$E_{snake} = \int_a^b [-\lambda(s)|\nabla I(v(s))|^2 + \alpha(s)\left(\frac{dv}{ds}\right)^2 + \beta(s)\left(\frac{d^2v}{ds^2}\right)^2] ds \quad (2.5)$$

Chapitre 02 Traitement des Images de mammographies

B.2. Résolution

Au cours de la résolution de l'équation de Snake, l'opérateur va minimiser l'énergie totale.

$$E(v(s)) = \int_{\Omega} \alpha(s)v'(s)^2 ds + \int_{\Omega} \beta(s)v''(s)^2 ds - \int_{\Omega} \lambda(s)|\nabla I(v(s))|^2 ds \quad (2.6)$$

v' et v'' présente respectivement les dérivées premières et secondes de v le long de la courbe et Ω la région, l'énergie à minimiser est donc donnée par formule (2.6) :

Pour minimiser cette énergie, on peut utiliser les équations d'Euler :

$$-\alpha v''(s) + \beta v^{(4)}(s) + \nabla P(v) = 0 \quad (2.7)$$

α , β et λ sont des coefficients constants.

Où

$$P(v) = -\lambda(s)|\nabla I(v(s))|^2 \quad (2.8)$$

$P(v)$ est l'énergie potentielle liée à l'image, afin de simplifier l'écriture, on pose

$$F(v) = -\nabla P(v) \quad (2.9)$$

L'équation de l'énergie à minimiser (avec α et β constants) devient donc :

$$-\alpha v''(s) + \beta v^{(4)}(s) = F(v(s)) \quad (2.10)$$

Les dérivées de l'équation de l'énergie sont résolues par la méthode de différences finie. On les met alors sous forme matricielle, nous donnant ainsi le schéma d'évolution suivant :

$$AV = F \quad (2.11)$$

Chapitre 02 Traitement des Images de mammographies

Avec A est une matrice à bande étroite dite penta-diagonale de taille $n \times n$ en fonction des coefficients α et β :

$$A = \begin{pmatrix} 2\alpha + 6\beta & -\alpha - 4\beta & \beta & 0 & 0 & \dots \\ -\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta & \beta & 0 & \dots \\ \beta & -\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta & \beta & \dots \\ 0 & \beta & -\alpha - 4\beta & 2\alpha + 6\beta & -\alpha - 4\beta & \dots \\ 0 & 0 & \beta & -\alpha - 4\beta & 2\alpha + 6\beta & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (2.12)$$

Ce schéma aboutit à l'équation :

$$(Id + \tau A)V^t = V^{t-1} + \tau F(V^{t-1}) \quad (2.13)$$

Soit :

$$V^t = (Id + \tau A)^{-1}(V^{t-1} + \tau F(V^{t-1})) \quad (2.14)$$

Où :

Id est la matrice identité de taille $(n \times n)$

τ est le pas du temps qui contrôle la vitesse de déplacement du Snake.

On déduit la position à l'itération t en fonction des forces liées à l'image et de la position $t-1$.

Le critère de convergence c'est lorsque V^t et V^{t-1} sont très proches, pour assurer l'arrêt du processus.

Chapitre 02 Traitement des Images de mammographies

C. L'algorithme du contour actif

L'algorithme ci-dessous montre le fonctionnement de l'équation de Snake :

Faire

Pour tous les points du Snake

Pour tous les points du voisinage

Calculer les énergies

Fin pour

Pour tous les points du voisinage

Normalisation

Fin pour

Minimiser pour obtenir le nouveau point

Fin pour

Jusqu'au critère d'arrêt

Cet algorithme est appliqué sur des images à niveau de gris et qui se résume dans l'organigramme présenté sur la figure 2.7.

Chapitre 02 Traitement des Images de mammographies

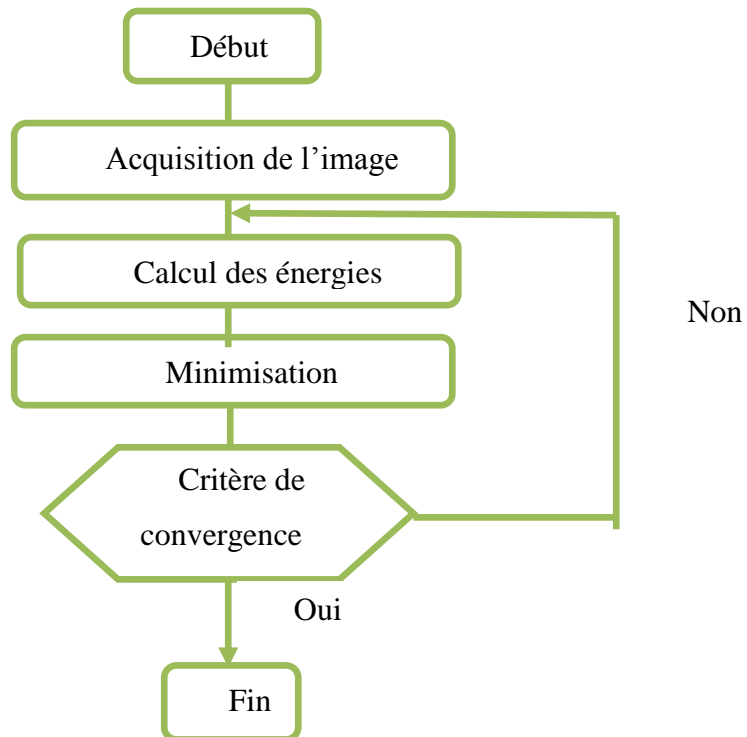


Figure 2.7 : Organigramme de la méthode de contour actif

Les exemples suivants (figure 2.8) sont une application de contours actifs (snakes) sur l'image mammographie en utilisant la base de données MIAS.

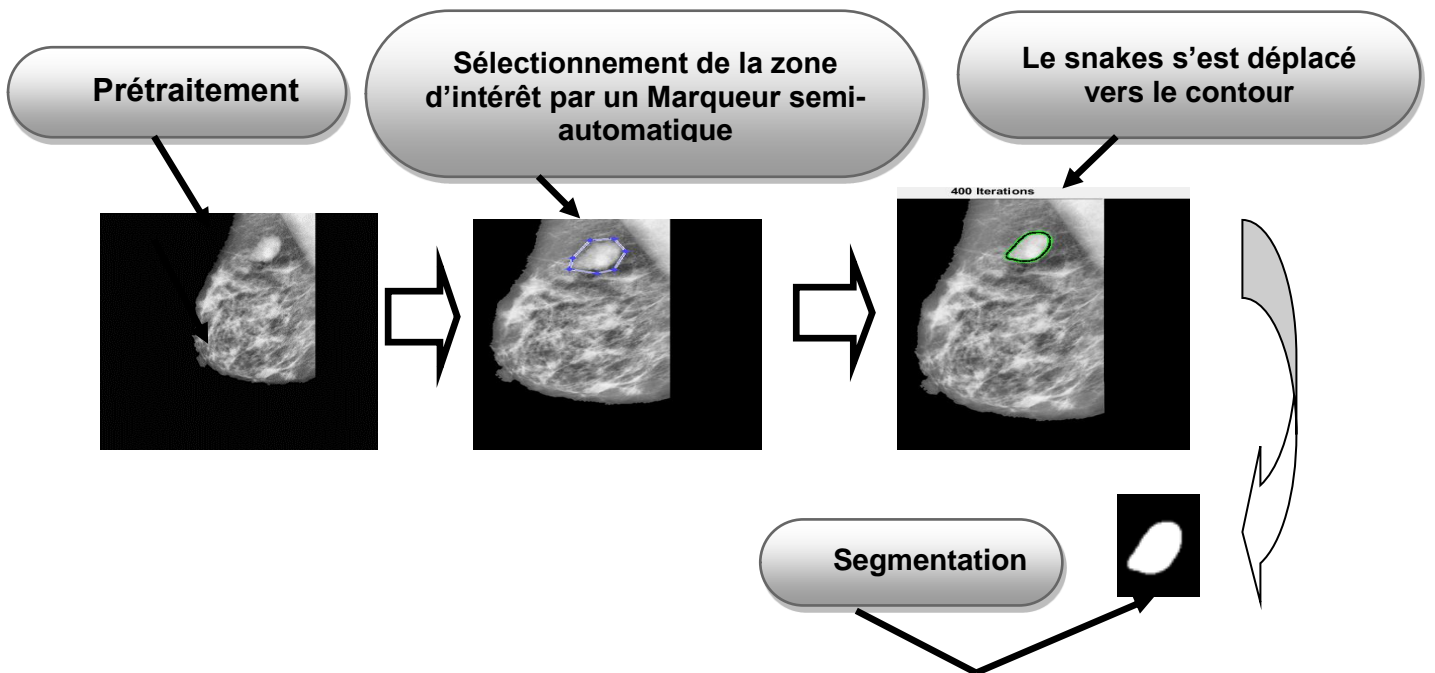


Figure 2.8: Exemple de segmentation de masse par contours actifs (snakes).

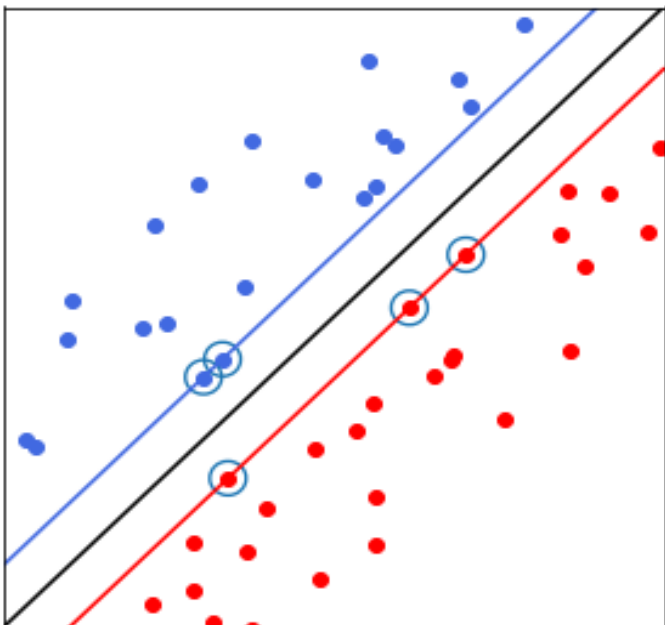
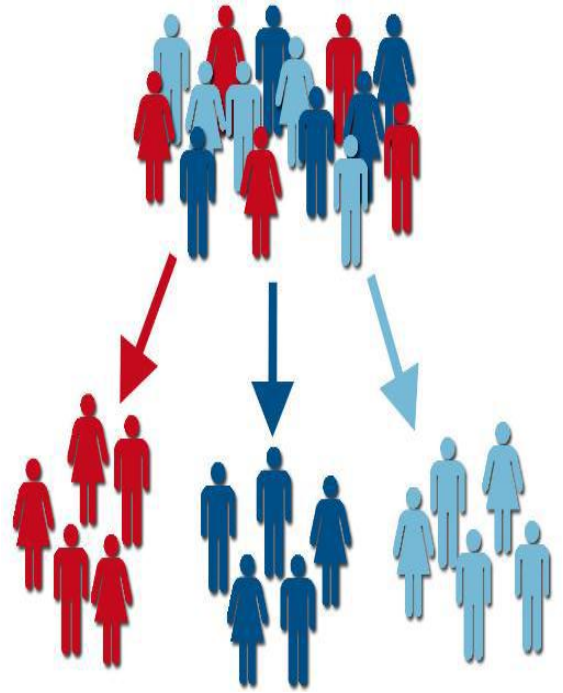
Chapitre 02 Traitement des Images de mammographies

2.4 Conclusion

Les étapes de prétraitement du processus de reconnaissance des formes sont le point de départ de toute autre analyse avancée impliquant des opérations de filtrage, d'amélioration du contraste et de segmentation.

Dans ce chapitre, nous avons présenté certaines des méthodes de prétraitement que nous avons utilisées dans notre projet.

Chapitre 03



Les machines à
Vecteur support
(SVM)

Chapitre 3

Les machines à vecteur support (SVM)

3.1. Introduction

Les machines à vecteur support se situent sur l'axe de développement de la recherche humaine des techniques d'apprentissage. Les SVMs sont une classe de techniques d'apprentissage introduite par Vladimir Vapnik au début des années 90 [34, 35], elles reposent sur une théorie mathématique solide à l'inverse des méthodes de réseaux de neurones. Elles ont été développées au sens inverse du développement des réseaux de neurones : ces derniers ont suivi un chemin heuristique de l'application et l'expérimentation vers la théorie ; alors que les SVMs sont venues de la théorie du son vers l'application. Les SVMs sont dans leur origine utilisées pour la classification binaire et la régression.

Elle repose sur l'existence d'un classifieur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonctions dites noyau (kernel) qui permettent une séparation optimale des données. Dans la présentation des principes de fonctionnements.

Aujourd'hui, elles sont utilisées dans différents domaines de recherche et d'ingénierie tel que le diagnostic médical, le marketing, la biologie, la reconnaissance de caractères manuscrits et de visages humains. Le présent chapitre introduit les machines à vecteur support, leurs origines théoriques, leurs différentes formes et leur optimisation. Le chapitre est organisé comme suit : il commence par introduire la relation entre l'apprentissage et les SVMs. Nous verrons ensuite la forme originale des SVMs : le cas binaire (le principe de fonctionnement).

Chapitre 3 Les machines à vecteur support (SVM)

3.2. Principe de fonctionnement

3.2.1. Apprentissage statistique

La théorie d'apprentissage statistique étudie les propriétés mathématiques des machines d'apprentissage [36]. Ces propriétés représentent les propriétés de la classe de fonctions ou modèles que peut implémenter la machine. L'apprentissage statistique utilise un nombre limité d'entrées (appelées exemples) d'un système avec les valeurs de leurs sorties pour apprendre une fonction qui décrit la relation fonctionnelle existante, mais non connue, entre les entrées et les sorties du système.

On suppose premièrement que les exemples d'apprentissage, appelés aussi exemples d'entraînement, sont générés selon une certaine probabilité inconnue (mais fixe) c'est-à-dire indépendants et identiquement distribués (iid). C'est une supposition standard dans la théorie d'apprentissage. Les exemples sont de dimension $m \in \mathbb{R}^m$ et dans le cas d'apprentissage supervisé, accompagnés d'étiquettes caractérisant leurs types ou classes d'appartenance. Si ces étiquettes sont dénombrables, on parle de classification sinon on parle de régression. Dans le cas d'une classification binaire cette étiquette est soit +1 ou -1. L'ensemble des exemples et leurs étiquettes correspondantes est appelé ensemble d'apprentissage.

Une machine efficace d'apprentissage est une machine qui apprend de l'ensemble d'entraînement une fonction qui minimise les erreurs de classification sur l'ensemble lui-même.

Soit $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ l'ensemble des exemples d'entraînement, avec $x_i \in \mathbb{R}^m$ et $y_i = \pm 1$, et soit $f(x)$ la fonction apprise par la machine d'apprentissage. On appelle le risque empirique $R_{emp}[f]$ le taux d'erreurs effectuées par la fonction f sur l'ensemble d'entraînement D .

$$R_{emp} = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) \quad (3.1)$$
$$L = \begin{cases} 1 & \text{si } y_i = f(x_i) \\ 0 & \text{sinon} \end{cases}$$

Chapitre 3 Les machines à vecteur support (SVM)

Puisque l'ensemble D ne représente qu'une simple partie de tout l'espace d'exemples, la fonction apprise f , qui minimise le risque empirique, peut se comporter mal avec les nouveaux exemples non vus à l'entraînement. C'est un phénomène très connu en apprentissage automatique appelé le sur-apprentissage ou apprentissage par cœur. Pour garantir que f , prenne en charge même les exemples non jamais vus, il faut contrôler sa capacité de généralisation mesurée souvent sur un autre ensemble d'exemples appelé ensemble de test réservé uniquement pour tester la machine apprise. La fonction f recherchée doit donc minimiser les erreurs de classification sur les deux ensembles d'entraînement et de test.

Trouver la fonction optimale f revient toujours à un problème d'optimisation, ce qui explique la forte relation entre l'apprentissage et l'optimisation. Avant de rechercher la fonction f , il faut définir son type puis rechercher ses paramètres.

Dans le cas des machines à vecteur support, la fonction recherchée est de forme linéaire. Les SVMs sont, donc, des systèmes d'apprentissage qui utilisent un espace d'hypothèses de fonctions linéaires dans un espace de caractéristique à haute dimension. Cette stratégie d'apprentissage introduite par Vapnik et ses co-auteurs [37] est une méthode très puissante. Elle a pu, en quelques années depuis sa proposition, conquérir la plupart des autres systèmes d'apprentissage dans une grande variété de domaines d'application.

3.3. SVM principe de fonctionnement général en premier

3.3.1. Notions de base : Hyperplan, marge et support vecteur

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classifieur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classifieur est un classifieur linéaire appelé hyperplan.

Dans le schéma qui suit, on détermine un hyperplan qui sépare les deux ensembles de points [38].

Chapitre 3 Les machines à vecteur support (SVM)

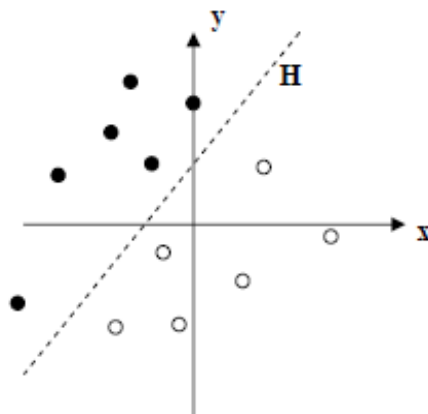


Figure 3.1 : Exemple d'un hyperplan séparateur [38].

3.3.2. SVMs binaires

Le cas le plus simple est celui où les données d'entraînement viennent uniquement de deux classes différentes (+1 ou -1), on parle alors de classification binaire. L'idée des SVMs est de rechercher un hyperplan (droite dans le cas de deux dimensions) qui sépare le mieux ces deux classes.

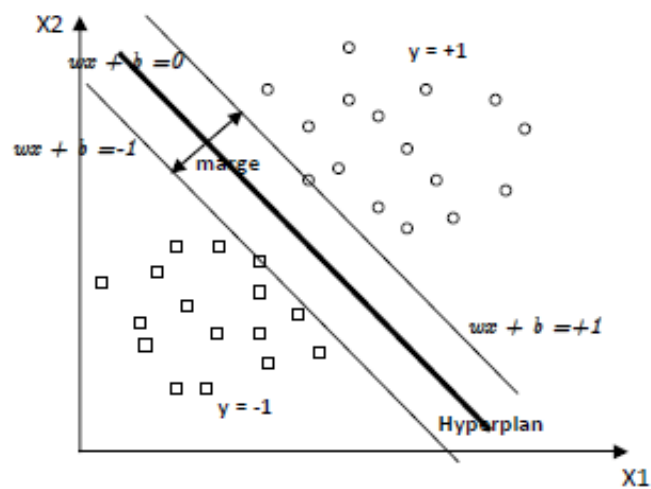


Figure 3.2 : SVM binaire

Si un tel hyperplan existe, c'est-à-dire si les données sont linéairement séparables, on parle d'une machine à vecteur support à marge dure (Hard margin).

Chapitre 3 Les machines à vecteur support (SVM)

3.3.2.1. SVM à marge dure :

L'hyperplan séparateur est représenté par l'équation (3.2) suivante :

$$f(x) = w^T X + b \quad (3.2)$$

Où w est un vecteur de T dimensions et b est un terme. La fonction de décision, pour un exemple x , peut être exprimée comme suit :

$$\begin{cases} \text{Classe} = 1 & \text{si } f(x) > 0 \\ \text{Classe} = -1 & \text{si } f(x) < 0 \end{cases} \quad (3.3)$$

Puisque les deux classes sont linéairement séparables, il n'existe aucun exemple qui se situe sur l'hyperplan, c.-à-d qui satisfait $f(x) = 0$. Il convient alors d'utiliser la fonction de décisions suivante :

$$\begin{cases} \text{Classe} = 1 & \text{si } f(x) > 1 \\ \text{Classe} = -1 & \text{si } f(x) < -1 \end{cases} \quad (3.4)$$

Les valeurs $+1$ et -1 à droite des inégalités peuvent être des constantes quelconques $+a$ et $-a$, mais en divisant les deux parties des inégalités par a , on trouve les inégalités précédentes qui sont équivalentes à l'équation (3.5)

$$y_i(w^T X_i + b) \geq 1, i = 1..n \quad (3.5)$$

L'hyperplan $w^T x + b = 0$ représente un hyperplan séparateur des deux classes, et la distance entre cet hyperplan et l'exemple le plus proche s'appelle la marge (figure 3.2). La région qui se trouve entre les deux hyperplans $w^T x + b = -1$ et $w^T x + b = +1$ est appelée la région de généralisation de la machine d'apprentissage. Plus cette région est importante, plus est la capacité de généralisation de la machine. La maximisation de cette région est l'objectif de la phase d'entraînement qui consiste, pour la méthode SVM, à rechercher l'hyperplan qui maximise la région de généralisation c.-à-d la marge. Un tel hyperplan est appelé "hyperplan de séparation optimale". En supposant que les données d'apprentissage ne contiennent pas des données bruitées (mal-étiquetées) et que les données de test suivent la même probabilité que celle des données d'entraînement, l'hyperplan de marge maximale va certainement maximiser la capacité de généralisation de la machine d'apprentissage.

Chapitre 3 Les machines à vecteur support (SVM)

La détermination de l'hyperplan optimal passe par la détermination de la distance euclidienne minimale entre l'hyperplan et l'exemple le plus proche des deux classes. Puisque le vecteur w est orthogonal sur l'hyperplan séparateur, la droite parallèle à w et reliant un exemple x à l'hyperplan est donnée par la formule (3.6) :

$$\frac{aw}{\|w\|} + x = 0 \quad (3.6)$$

Où a représente la distance entre x et l'hyperplan. La résolution de (3.6), donne (3.7) :

$$a = -\frac{w^T x + b}{\|w\|} \quad (3.7)$$

La distance de tout exemple de l'hyperplan doit être supérieure ou égale à la marge δ :

$$\frac{y_i(w^T x_i + b)}{\|w\|} \geq \delta \quad (3.8)$$

Si une paire (w, b) est une solution alors (aw, ab) est une solution aussi où a est un scalaire. On impose alors la contrainte (3.9) suivante :

$$\|w\|\delta \geq 1 \quad (3.9)$$

Pour trouver l'hyperplan séparateur qui maximise la marge, on doit déterminer, à partir des deux dernières inégalités, le vecteur w qui possède la norme euclidienne minimale et qui vérifie la contrainte de l'équation (3.5), de bonne classification des exemples d'entraînement. L'hyperplan séparateur optimal peut être obtenu en résolvant le problème de l'équation (3.10) :

$$\left\{ \begin{array}{l} \text{Minimiser} \quad \frac{1}{2} \|w\|^2 \\ \text{sous contraintes} \\ y_i(w^T x_i + b) \geq 1 \quad \forall i = 1..n \end{array} \right. \quad (3.10)$$

Chapitre 3 Les machines à vecteur support (SVM)

Remarquons que nous pouvons obtenir le même hyperplan même en supprimant toutes les données qui vérifient l'inégalité de la contrainte. Les données qui vérifient l'égalité de la contrainte s'appellent les vecteurs supports, et ce sont ces données seules qui contribuent à la détermination de l'hyperplan. Dans la figure 3.2, les données qui se trouvent sur les deux droites +1 et -1 représentent les vecteurs supports.

Le problème de l'équation (3.10) est un problème de programmation quadratique avec contraintes linéaires. Dans ce problème, les variables sont w et b , c.-à-d que le nombre de variables est égal à $m + 1$. Généralement, le nombre de variables est important ce qui ne permet pas d'utiliser les techniques classiques de programmation quadratique. Dans ce cas le problème (3.10) est convertit en un problème dual équivalent sans contraintes de l'équation (3.11) qui introduit les multiplicateurs de Lagrange :

$$Q(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i \{y_i(w^T x_i + b) - 1\} \quad (3.11)$$

Où les α_i sont les multiplicateurs non négatifs de Lagrange. L'optimum de la fonction objective Q peut être obtenu en la minimisant par rapport à w et b et en la maximisant par rapport aux α_i .

À l'optimum de la fonction objective, ses dérivées par rapports aux variables w et b s'annulent ainsi que le produit des α_i aux contraintes (équation 3.12) :

$$\begin{cases} \frac{\partial Q(w, b, \alpha)}{\partial w} = 0 & (a) \\ \frac{\partial Q(w, b, \alpha)}{\partial b} = 0 & (b) \\ \alpha_i \{y_i(w^T x_i + b) - 1\} = 0 & (c) \\ \alpha_i \geq 0 & (d) \end{cases} \quad (3.12)$$

Chapitre 3 Les machines à vecteur support (SVM)

De (3.12.a) on déduit :

$$\begin{cases} w = \sum_{i=1}^n \alpha_i y_i x_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (3.13)$$

En remplaçant dans 3.11, on obtient le problème dual à maximiser suivant :

$$\begin{cases} \text{Maximiser} & Q(\alpha) = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j \\ \text{sous contraintes} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \end{cases} \quad (3.14)$$

Si le problème de classification est linéairement séparable, une solution optimale pour les α_i existe. Les exemples ayant des $\alpha_i \neq 0$ représentent les vecteurs supports appartenant aux deux classes. La fonction de décision est donnée par :

$$f(x) = \sum_S \alpha_i y_i x^T x_i + b \quad (3.15)$$

Où S représente l'ensemble des vecteurs supports. b peut être calculé à partir de n'importe quel vecteur support par l'équation (3.16) :

$$b = y_i - w^T x_i \quad (3.16)$$

D'un point de vue précision, on prend la moyenne de b pour tous les vecteurs supports :

$$b = \frac{1}{|S|} \sum_{i \in S} y_i - w^T x_i \quad (3.17)$$

La fonction de décision f peut être calculée, donc, pour chaque nouvel exemple x par l'équation (3.15) et la décision peut être prise comme suit :

Chapitre 3 Les machines à vecteur support (SVM)

$$\begin{cases} x \in \text{Classe} + 1 & \text{si } f(x) > 0 \\ x \in \text{Classe} - 1 & \text{si } f(x) < 0 \\ x & \text{inclassifiable si } f(x) = 0 \end{cases} \quad (3.18)$$

La zone $-1 < f(x) < 1$ est appelée la zone de généralisation.

Si on prend un exemple x_k de l'ensemble d'entraînement appartenant à la classe y_k et on calcule sa fonction de décision $f(x_k)$, on peut se trouver dans l'un des cas suivants :

1. $y_k * f(x_k) > 1$: dans ce cas l'exemple est bien classé et ne se situe pas dans la zone de la marge. Il ne représente pas un vecteur support.
2. $y_k * f(x_k) = 1$: dans ce cas l'exemple est bien classé et se situe aux frontières de la zone de la marge. Il représente un vecteur support.
3. $0 < y_k * f(x_k) < 1$: dans ce cas l'exemple est bien classé et se situe dans la zone de la marge. Il ne représente pas un vecteur support.
4. $y_k * f(x_k) < 0$ Dans ce cas l'exemple se situe dans le mauvais coté, il est mal classé et ne représente pas un vecteur support.

3.3.2.2. SVM à marge souple

En réalité, un hyperplan séparateur n'existe pas toujours, et même s'il existe, il ne représente pas généralement la meilleure solution pour la classification. En plus une erreur d'étiquetage dans les données d'entraînement (un exemple étiqueté +1 au lieu de -1 par exemple) affectera crucialement l'hyperplan.

Dans le cas où les données ne sont pas linéairement séparables, ou contiennent du bruit (outliers : données mal étiquetées) les contraintes de l'équation (3.5) ne peuvent être vérifiées, et il y a nécessité de les relaxer un peu. Ceci peut être fait en admettant une certaine erreur de classification des données (figure 3.3) ce qui est appelé "SVM à marge souple (Soft Margin)".

Chapitre 3 Les machines à vecteur support (SVM)

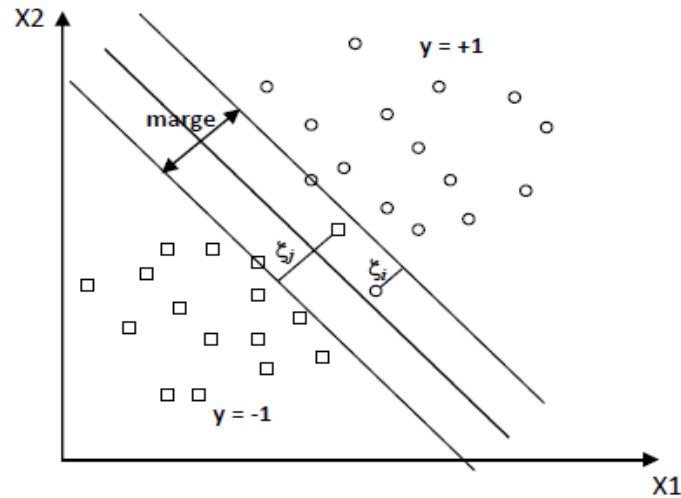


Figure 3.3 : SVM binaire à marge souple

On introduit alors sur les contraintes des variables ε_i dites de relaxation pour obtenir la contrainte de l'équation (3.19) :

$$y_i(w^T x_i + b) \geq 1 - \varepsilon_i, i = 1..n \quad (3.19)$$

Grâce aux variables de relaxation non négatives ε_i , un hyperplan séparateur existera toujours. Si $\varepsilon_i < 1$, x_i ne respecte pas la marge mais reste bien classé, sinon x_i est mal classé par l'hyperplan. Dans ce cas, au lieu de rechercher uniquement un hyperplan séparateur qui maximise la marge, on recherche un hyperplan qui minimise aussi la somme des erreurs permises c-à-d minimisé $Q(w) = \sum_{i=1}^n \varepsilon_i$.

En combinant avec l'équation (3.10), on obtient le problème primal de l'équation (3.20) Suivant :

$$\left\{ \begin{array}{l} \text{Minimiser} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \\ \text{sous contraintes} \\ y_i(w^T x_i + b) \geq 1 - \varepsilon_i, i = 1..n \\ \varepsilon_i \geq 0 \end{array} \right. \quad (3.20)$$

Où C est un paramètre positif libre (mais fixe) qui représente une balance entre les deux termes de la fonction objective (la marge et les erreurs permises) c.-à-d entre la maximisation de la marge et la minimisation de l'erreur de classification. On

Chapitre 3 Les machines à vecteur support (SVM)

obtient le problème dual de l'équation (3.21) où on introduit les multiplicateurs de Lagrange α_i et β_i :

$$Q(\omega, b, \alpha, \varepsilon, \beta) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \varepsilon_i - \sum_{i=1}^n \alpha_i y_i (\omega^T x_i + b) - 1 + \varepsilon_i - \sum_{i=1}^n \beta_i \varepsilon_i \quad (3.21)$$

À la solution optimale, les dérivées par rapport aux variables ω, b, α, β s'annulent ainsi que le produit des contraintes aux multiplicateurs. Les conditions suivantes sont alors vérifiées :

$$\begin{cases} \frac{\partial Q(\omega, b, \varepsilon, \alpha, \beta)}{\partial \omega} = 0 & (a) \\ \frac{\partial Q(\omega, b, \varepsilon, \alpha, \beta)}{\partial b} = 0 & (b) \\ \alpha_i \{y_i (\omega^T x_i + b) - 1 + \varepsilon_i\} = 0 & (c) \\ \beta_i \varepsilon_i = 0 & (d) \\ \alpha_i \geq 0, \beta_i \geq 0, \varepsilon_i \geq 0 & (e) \end{cases} \quad (3.22)$$

On déduit :

$$\begin{cases} \omega = \sum_{i=0}^n \alpha_i y_i x_i \\ \sum_{i=0}^n \alpha_i y_i = 0 \\ \alpha_i + \beta_i = 0 \end{cases} \quad (3.23)$$

En remplaçant l'équation (3.23) dans l'équation (3.21), on obtient le problème dual (3.24) suivant :

$$\begin{cases} \text{Minimiser} & Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{sous contraintes} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{cases} \quad (3.24)$$

Chapitre 3 Les machines à vecteur support (SVM)

La seule différence avec la SVM à marge dure est que les α_i ne peuvent pas dépasser C , ils peuvent être dans l'un des trois cas suivants :

1. $\alpha_i = 0 \rightarrow \beta_i = C \rightarrow \varepsilon_i = 0$: x_i est bien classé,
2. $0 < \alpha_i < C \rightarrow \beta_i > 0 \rightarrow \varepsilon_i = 0 \rightarrow y_i(\omega^T x_i + b) = 1$: x_i est un vecteur support et est appelé dans ce cas vecteur support non borné (unbounded),
3. $\alpha_i = C \rightarrow \beta_i = 0 \rightarrow \varepsilon_i \geq 0 \rightarrow y_i(\omega^T x_i + b) = 1 - \varepsilon_i$: x_i est un vecteur support appelé dans ce cas vecteur support borné (bounded). Si $0 < \varepsilon_i < C$, x_i est bien classé, sinon x_i est mal classé.

Ces conditions sur les α_i sont appelées les conditions de Karush-Kuhn-Tucker (KKT) [39, 40], elles sont très utilisées par les algorithmes d'optimisation pour rechercher les α_i optimaux et par conséquent l'hyperplan optimal.

La fonction de décision est alors calculée de la même manière que dans le cas des SVMs à marge dure mais uniquement à base des vecteurs supports non bornés par :

$$f(x) = \sum_{i \in U} \alpha_i y_i x_i^T x + b \quad (3.25)$$

Pour les vecteurs supports non bornés, nous avons :

$$b = y_i - \omega^T x_i \quad (3.26)$$

Pour garantir une bonne précision, on prend la moyenne de b pour tous les vecteurs supports non bornés :

$$b = \frac{1}{|U|} \sum_{i \in U} y_i - \omega^T x_i \quad (3.27)$$

3.3.2.3. Utilisation des noyaux (SVM non linéaire)

Le fait d'admettre la mal-classification de certains exemples, ne peut pas toujours donner une bonne généralisation pour un hyperplan même si ce dernier est optimisé (figure 3.4).

Chapitre 3 Les machines à vecteur support (SVM)

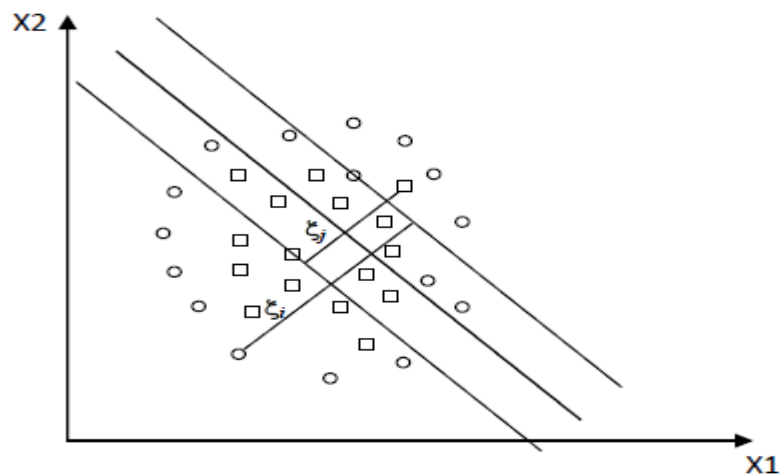


Figure 3.4 : Mal-adaptation de l'hyperplan aux problèmes réels de classification.

Plutôt qu'une droite, la représentation idéale de la fonction de décision serait une représentation qui colle le mieux aux données d'entraînement (figure 3.5).

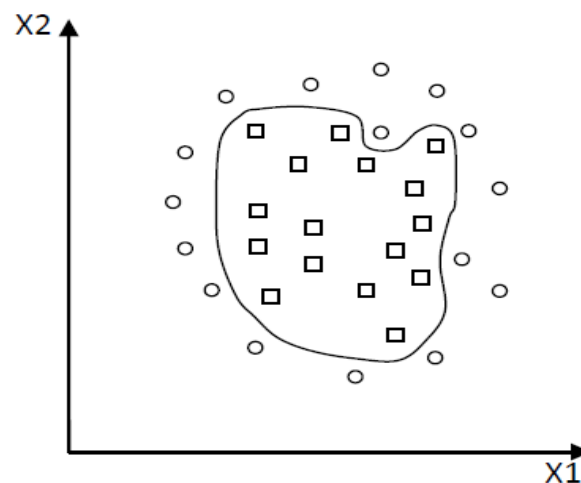


Figure 3.5: Représentation idéale de la fonction de décision.

La détermination d'une telle fonction non linéaire est très difficile voire impossible. Pour cela les données sont amenées dans un espace où cette fonction devient linéaire (figure 3.6), cette astuce permet de garder les mêmes modèles de problèmes d'optimisation vus dans les sections précédentes, utilisant les SVMs basées essentiellement sur le principe de séparation linéaire. Cette transformation d'espace est réalisée souvent à l'aide d'une fonction $F = \{\phi(x) | x \in X\}$ appelé

Chapitre 3 Les machines à vecteur support (SVM)

"Mapping function" et le nouvel espace est appelé espace de caractéristiques "Features space".

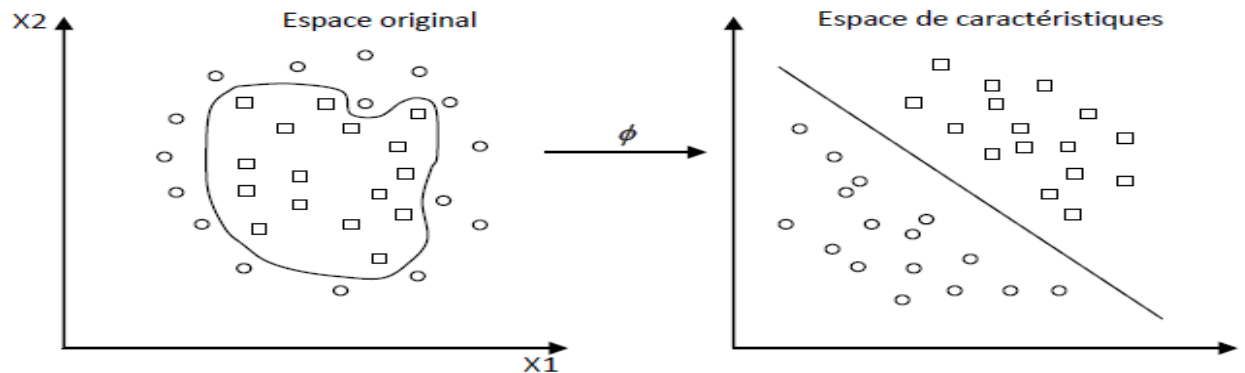


Figure 3.6 : Transformation d'espace

Dans ce nouvel espace de caractéristiques, la fonction objective à optimiser de l'équation (3.24) devient :

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \quad (3.28)$$

Où $\langle \phi(x_i), \phi(x_j) \rangle$ est le produit scalaire des deux images des vecteurs x_i et x_j dans le nouvel espace et dont le résultat est un scalaire.

Dans le calcul de l'optimum de la fonction (3.28), on utilise une astuce appelée "Noyau" ("Kernel"), au lieu de calculer $\phi(x_i), \phi(x_j)$ et leur produit scalaire, on calcule plutôt une fonction $K(x_i, x_j)$ qui représente à la fois les deux transformations (qui peuvent être inconnues) et leur produit scalaire. Cette fonction permet de surmonter le problème de détermination de la transformation ϕ et permet d'apprendre des relations non linéaires par des machines linéaires. La fonction $K(x_i, x_j)$ peut être vue comme une matrice $G[n, n]$ dite de Gram [41] qui représente les distances entre tous les exemples :

$$\begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{bmatrix}$$

Chapitre 3 Les machines à vecteur support (SVM)

Pour qu'une matrice G (fonction K) soit un noyau, il faut qu'elle respecte les conditions de Mercer [42,43], c.-à-d qu'elle doit être semi-définie positive (symétrique et n'a pas de valeurs propres négatives). La construction de tels noyaux est largement étudiée dans [44,45], néanmoins, il existe certains noyaux qui sont très utilisés et qui sont considérés comme standards 3.3.2.3. Une fois le noyau choisi, la fonction objective 3.28 peut être calculée comme suit 3.29:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.29)$$

Et la fonction de décision devient :

$$f(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x_j) + b \quad (3.30)$$

Où S représente l'ensemble des vecteurs supports.

Exemples de noyaux

- Noyau linéaire : Si les données sont linéairement séparables, on n'a pas besoin de changer d'espace, et le produit scalaire suffit pour définir la fonction de décision :

$$K(x_i, x_j) = x_i^T x_j \quad (3.31)$$

- Noyau polynomial : Le noyau polynomial élève le produit scalaire à une puissance naturelle d :

$$K(x_i, x_j) = (x_i^T x_j)^d \quad (3.32)$$

Si $d = 1$ le noyau devient linéaire. Le noyau polynomial dit non homogène $K(x_i, x_j) = (x_i^T x_j + C)^d$ est aussi utilisé.

Chapitre 3 Les machines à vecteur support (SVM)

- Noyau *RBF* : Les noyaux *RBF* (Radial Basis functions) sont des noyaux qui peuvent être écrits sous la forme : $K(x_i, x_j) = f(d(x_i, x_j))$ où d est une métrique sur X et f est une fonction dans R . Un exemple des noyaux *RBF* est le noyau Gaussien (3.33) :

$$K(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \quad (3.33)$$

Où σ est un réel positif qui représente la largeur de bande du noyau.

Exemple de séparation non linéaire

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau.

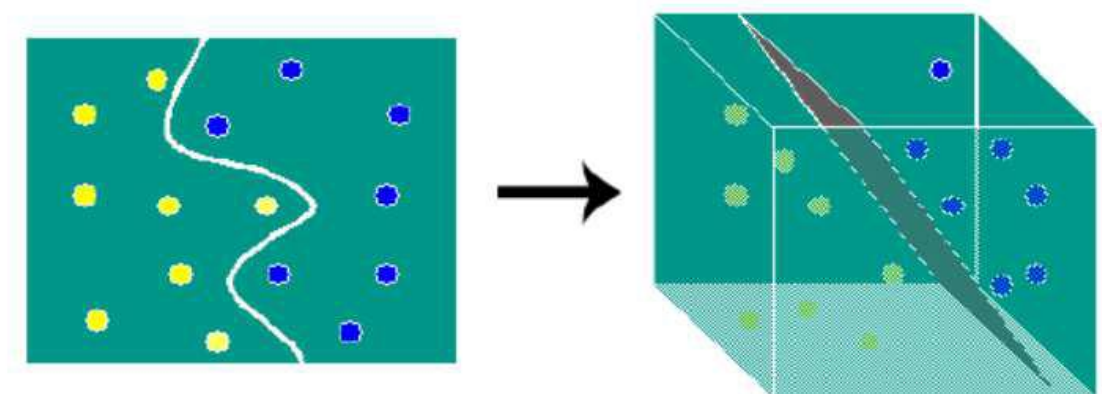


Figure 3.7 : Exemple de changement de l'espace de données.

Chapitre 3 Les machines à vecteur support (SVM)

3.4. Conclusion

Dans ce chapitre, nous avons tenté de présenter de manière simple et complète le concept de système d'apprentissage introduit par Vladimir Vapnik, les Machines à Vecteurs de Support. Nous avons donné une vision générale et une vision purement mathématiques des SVM.

Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données. Nous avons exposé les cas linéairement séparables et les cas non linéairement séparables qui nécessitent l'utilisation de fonction noyau (kernel) pour changer d'espace. Cette méthode est applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multi classe.

Donc nous avons conclu que les machines à vecteurs de support sont un moyen intéressant de classer les données.

Chapitre 04



Conception et Implémentation D'un système D'aide au Diagnostic assisté Par ordinateur (CAOx)



Chapitre 04

Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

4.1. Problématique

En raison de l'augmentation du nombre de mammographies réalisées ces dernières années, il est encore difficile pour les radiologues de fournir une analyse précise et homogène, d'où l'importance des aides au diagnostic développées au cours des deux dernières décennies. Un système d'aide à la décision (CAOx) dédié à l'imagerie médicale est un programme numérique qui consiste à assister les experts dans l'interprétation et la classification des images médicales. Son rôle est de détecter une ou plusieurs tumeurs sur une image et de les classer comme malignes ou bénignes.

L'objectif de ce projet est de développer un système semi-automatique d'aide à la détection et à la classification des masses (opacité) dans les images mammographiques à l'aide d'actif contour et de SVM.

4.2. Base de données

Nous utilisons la base de données [46] Mini-MIAS, qui est une version simplifiée de la base de données d'images originale MIAS. Les images sont de type MLO (Lateral Oblique Incidence) et chaque image à une résolution spatiale de 1024 x 1024 pixels. Les différentes pathologies mammaires de la base de données MIAS, objet de notre étude, représentent principalement des masses rondes, des spéculums et des distorsions architecturales bénignes et malignes. Notre base de données contient 56 images réparties comme suit :

- (21) images de masses malignes.
- (35) images de masses bénignes.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

4.3. Conception du système d'aide au diagnostic assisté par ordinateur

La conception du système d'aide au diagnostic proposé nécessite quatre étapes :

1. Le prétraitement des images mammographiques.
2. La segmentation
3. Détermination des descripteurs des formes de masses
4. La classification des masses en tant que malignes ou bénignes.

Pour cela nous allons suivre la démarche présentée dans la figure 4.1.

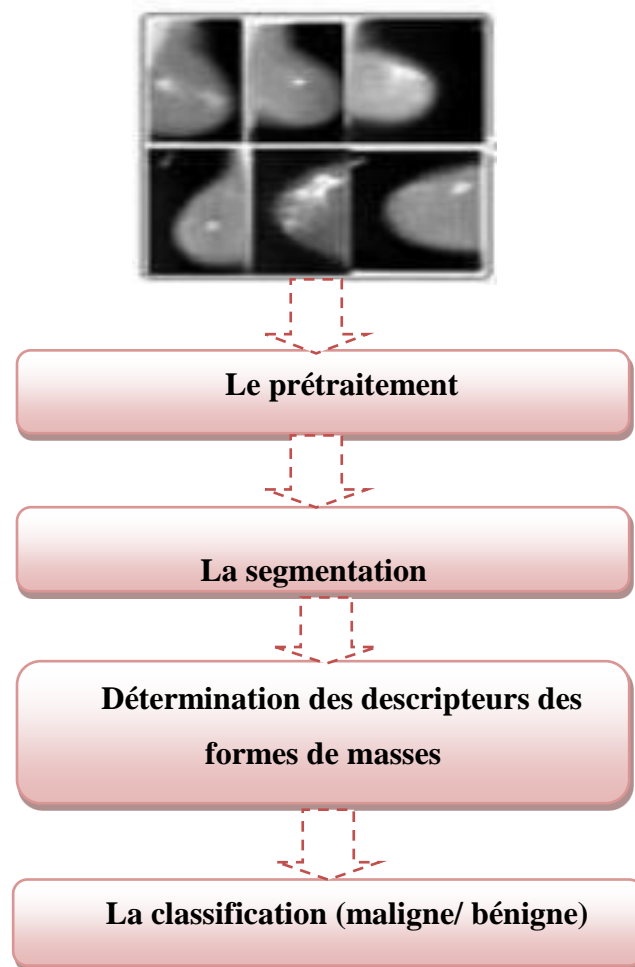


Figure 4.1: les étapes proposées de la conception du système d'aide au diagnostic

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Dans ce qui suit, nous allons détailler ces différentes étapes comme indiqué dans la figure 4.2.

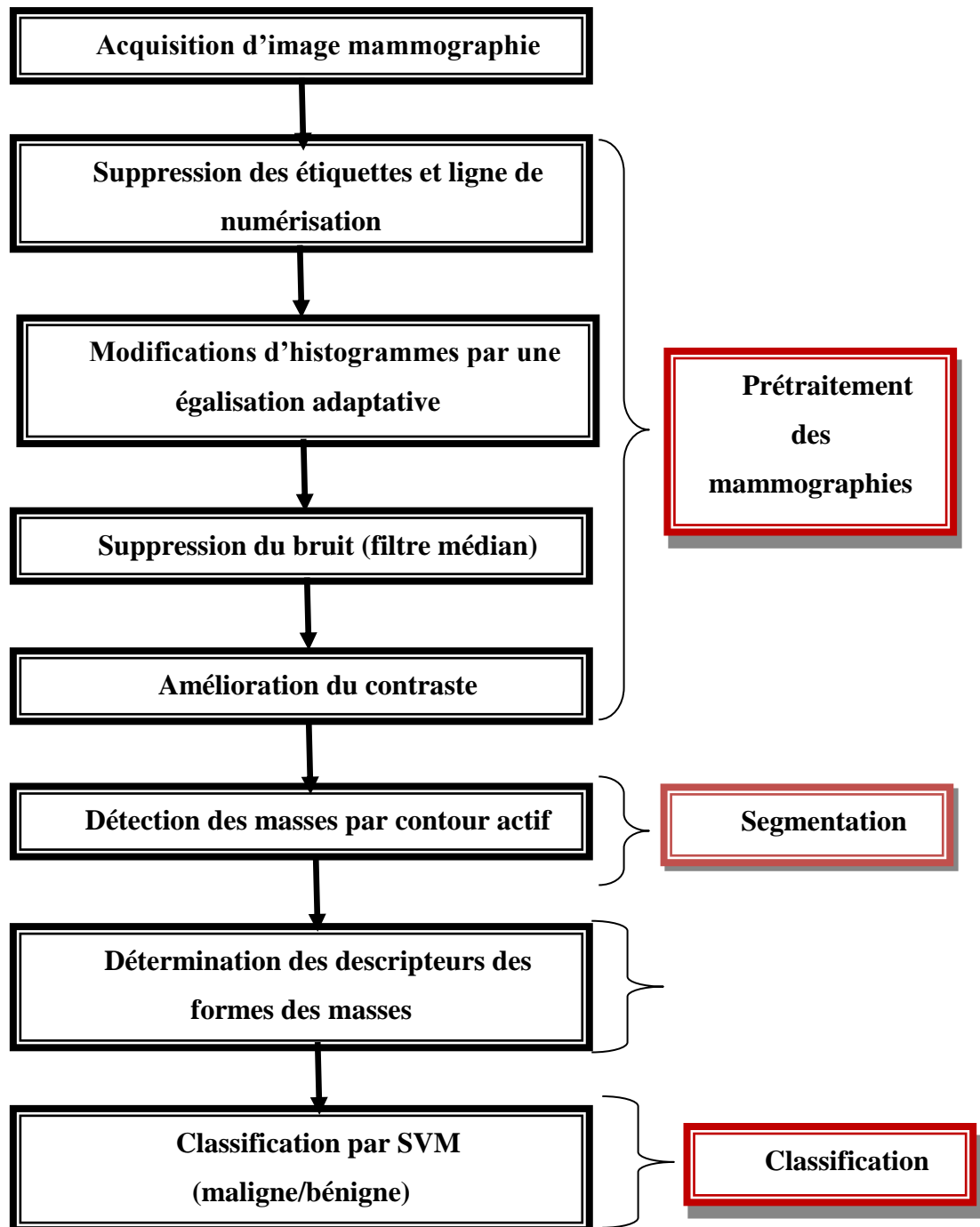


Figure 4.2 : Les différentes étapes du traitement des mammographies.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

4.3.1. Prétraitement des données

Les images acquises, qu'elles soient médicales ou non, sont souvent contaminées par des bruits de capteur et de lumière. Pour cela, chaque analyse d'image nécessite une étape de prétraitement. Dans notre travail, nous avons utilisé un filtre médian.

A. Suppression des lignes et des étiquettes de numérisation

Nous avons effectué la détection et la suppression des lignes de balayage à l'aide de deux fonctions : Seuillage et suppression des petites régions. Nous avons supprimé automatiquement toutes les zones qui ne font pas partie du sein sur la mammographie d'origine. La figure 4.3 illustre un exemple tiré de la base de données mini-Mias

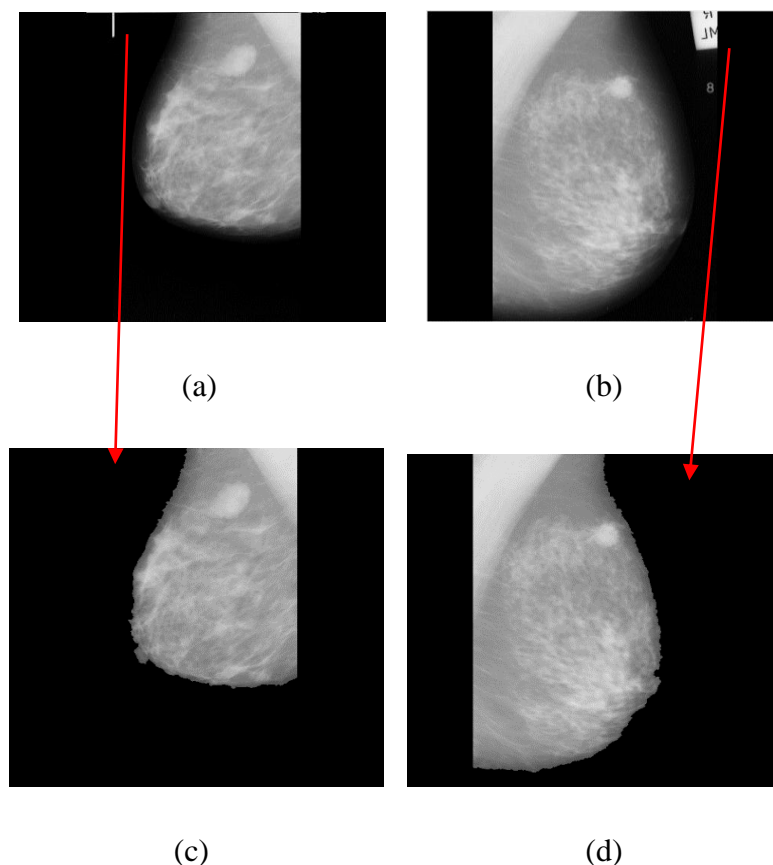


Figure 4.3 : Suppression des artefacts : (a) et (b) images originelles, (c) et (d) après élimination des étiquettes.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

A.1. Amélioration de la qualité d'images par l'égalisation adaptative d'histogramme

Nous avons amélioré le contraste en utilisant l'égalisation adaptative d'histogramme avec une implémentation sous Matlab. (Figure4.4) :

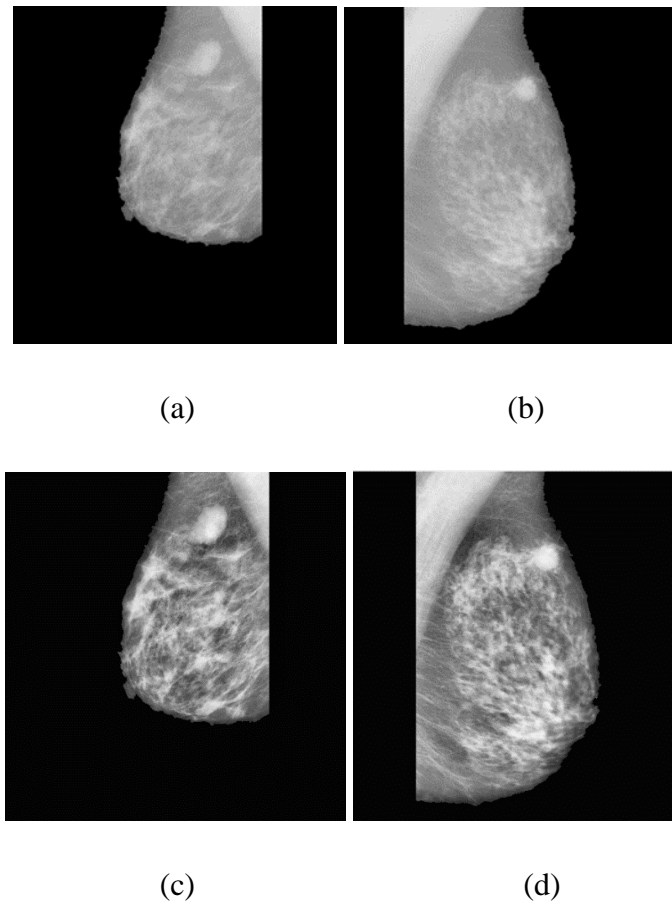


Figure 4.4 : Amélioration du contraste. (a) et (b) les images après la suppression des raies et des étiquettes, (c) et (d) amélioration de la qualité d'image (contraste) par égalisation adaptative d'histogramme.

B. Filtrage

Nous avons utilisé un filtre médian pour éliminer le bruit. Les résultats obtenus sont présentés dans la figure 4.5.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

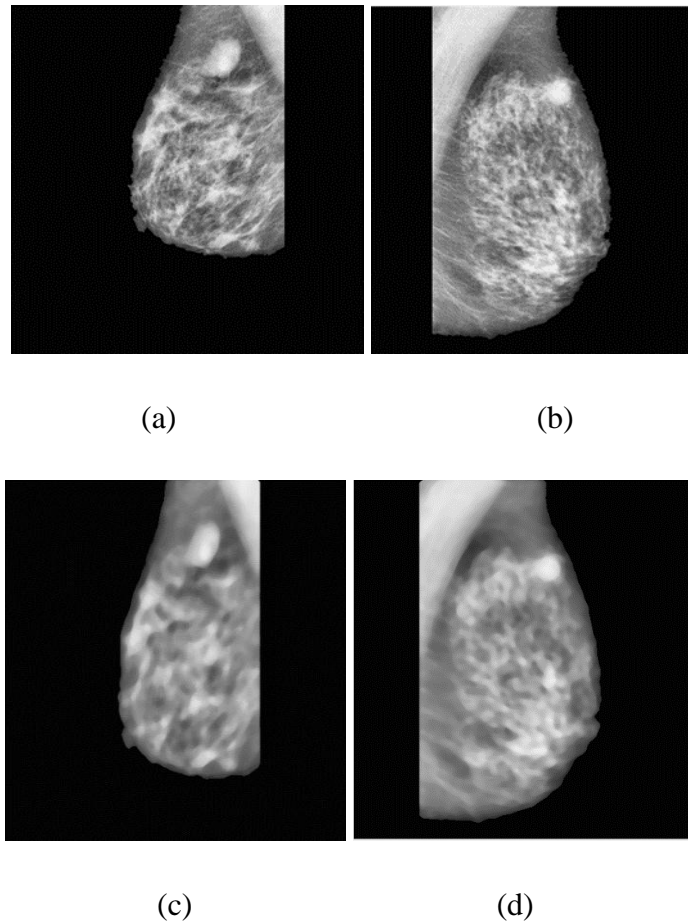


Figure 4.5 : Suppression du bruit : (a) et (b) images obtenues précédemment, (c) et (d) images filtrées.

C. Rehaussement de contraste

Après le filtrage de l'image, un rehaussement de contraste est effectué pour mettre en évidence toute les taches de haute intensité, autrement toutes les régions susceptibles d'être des masses. Dans un premier temps, nous avons amélioré le contraste en utilisant un filtre morphologique (voir chapitre2). Cette ouverture va éliminer les petites composantes et de séparer les grosses particules aux endroits où elles sont plus fines.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

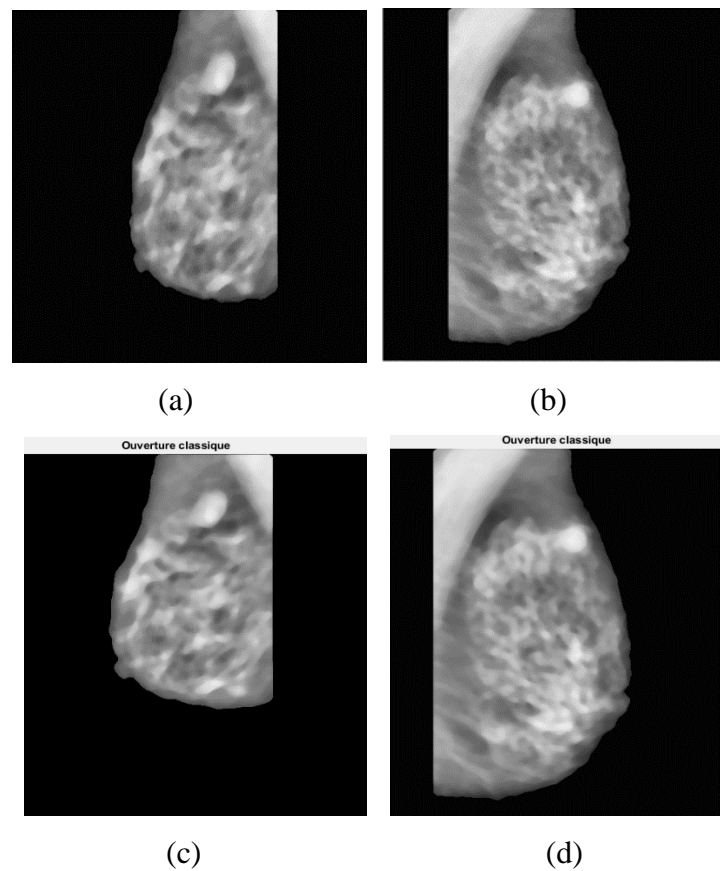


Figure 4.6: Amélioration du contraste. (a) et (b) les images filtrées, (c) et (d) rehaussement de contraste par une ouverture.

- **Filtre alternés séquentiels FAS**

Pour améliorer encore plus le contraste, nous avons décidé d'appliquer le FAS (Filtre alternés séquentiels) (voir chapitre2). L'algorithme proposé dans cette étape consiste à améliorer le contraste de l'image en utilisant des filtres de type morphologiques élémentaires, tels que les ouvertures et les fermetures qui permettent de réduire la complexité d'une image sans dégrader les contrastes les plus Importants. (Figure4.7) :

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

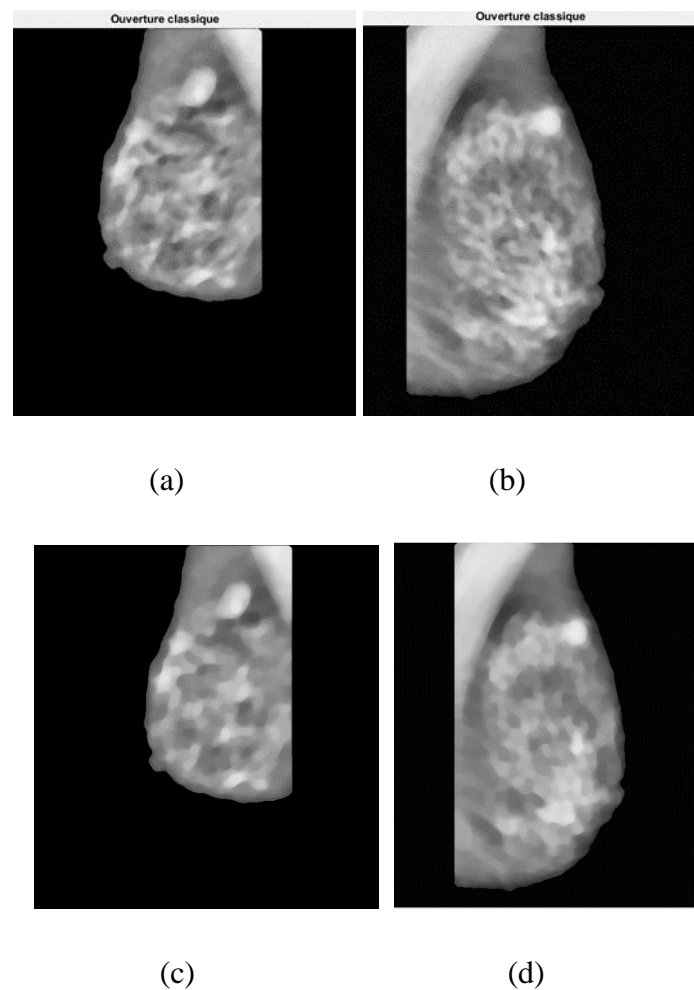


Figure 4.7 : Amélioration du contraste : (a) et (b) images filtrées par ouverture, (c) et (d) images avec rehaussement de contraste par filtre FAS.

4.3.2. Segmentation

Une fois l'image filtrée et rehaussée, une segmentation est nécessaire pour extraire la masse de l'image. Dans notre travail, nous avons utilisé la méthode de contours actifs afin de délimiter la pathologie cible et de la classer par la suite comme étant bénigne ou maline.

❖ Contour actif

La méthode de contours actifs (voir chapitre 2) consiste à positionner dans l'image, plus précisément au voisinage de la forme à détecter, un contour initial.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Pour cela, nous avons proposé un marquage semi-automatique pour sélection la zone d'intérêt (contour initial).

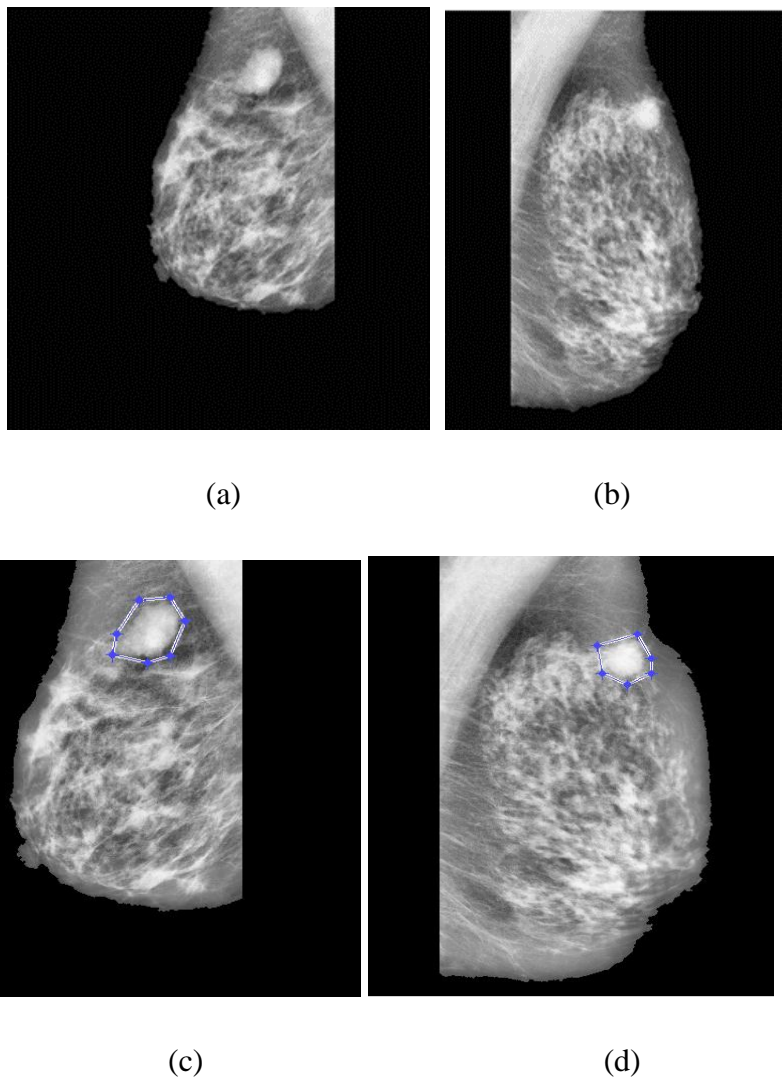


Figure 4.8 : (a)et (b) sont des images prétraitées, (c) et (d) sélectionnement de la zone d'intérêt par un marqueur semi-automatique

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Ce contour initial va subir une déformation sous l'effet de plusieurs forces (voir chapitre 2) :

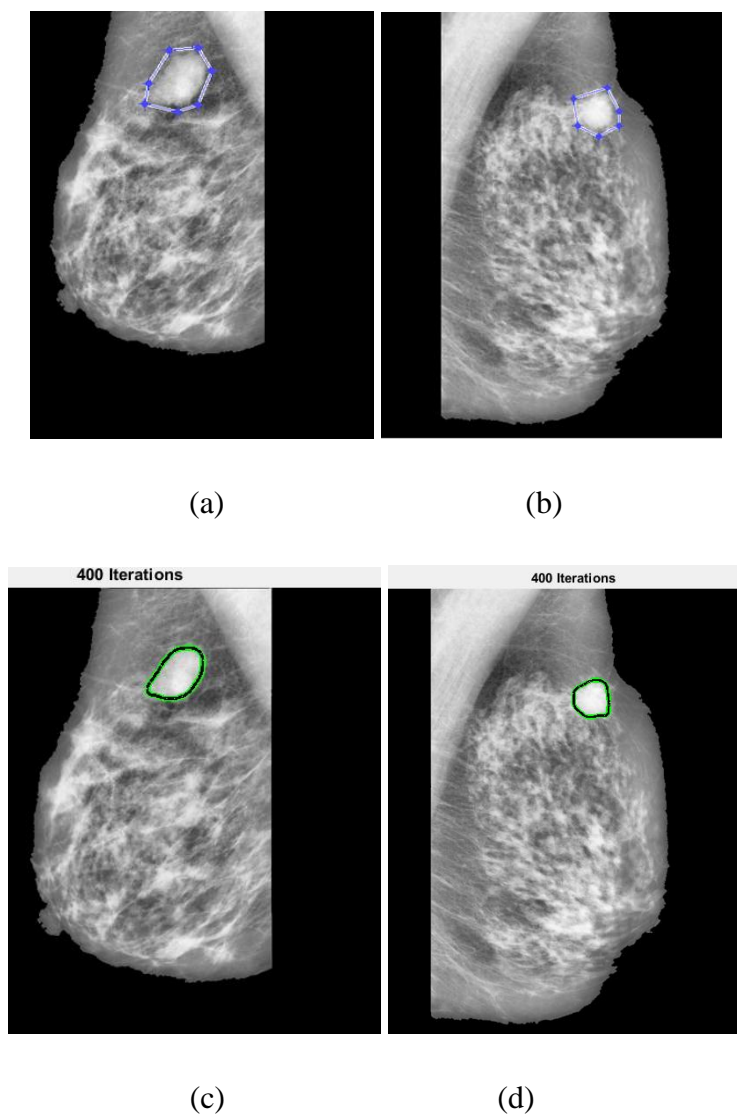
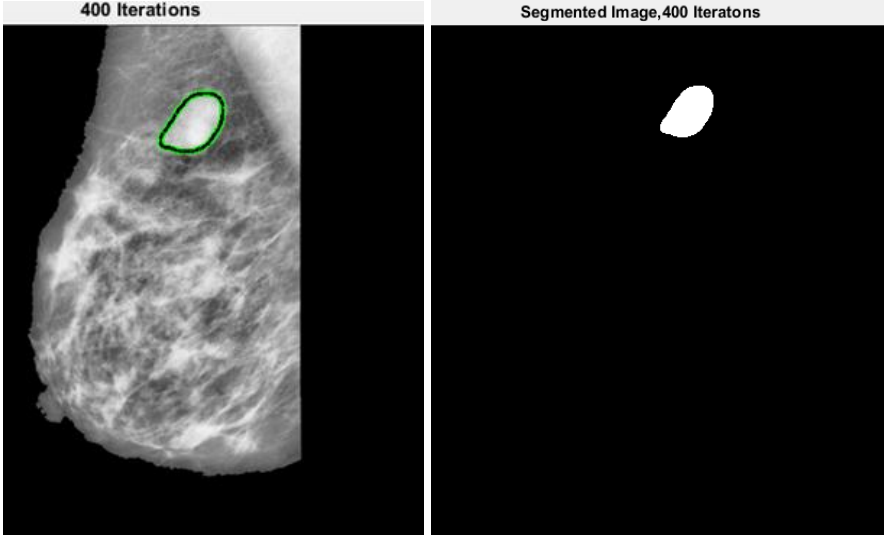


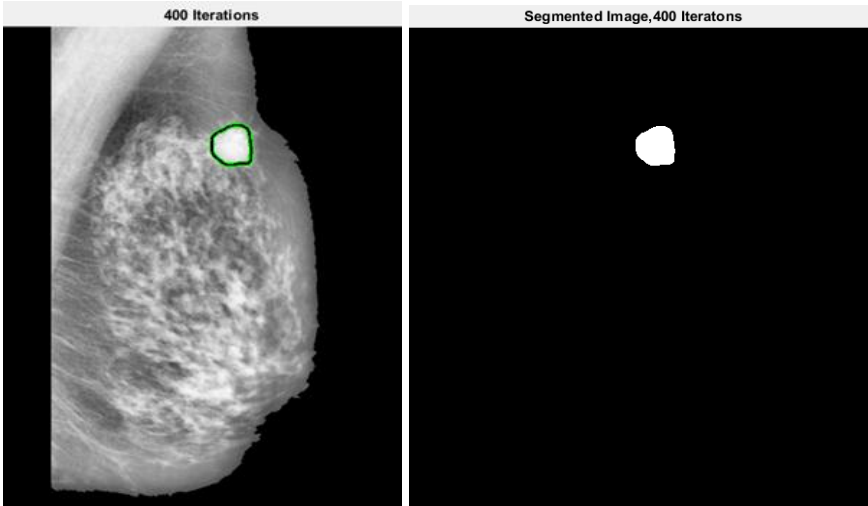
Figure 4.9: (a) et (b) des images avec un contour initial, (c) et (d) avec contour actif (snake)

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Résultat de la segmentation par contour actif de l'image (a) mdb015 et (b) mdb202.



(a)



(b)

Figure 4.10 : Résultat de la segmentation par contour actif des deux images (a) mdb015 et (b) mdb202.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

4.3.3. Descripteurs de forme

L'analyse d'images mammographiques permet non seulement de détecter les lésions mammaires, mais également de déterminer si elles sont de nature maligne ou bénigne en se basant sur des caractéristiques extraites de la masse (souvent des propriétés structurelles, photométriques et morphologiques). Dans cette étude, nous nous intéressons aux caractéristiques morphologiques (descripteurs) en termes de géométrie, telles que : Area (surface) et Périmètre

A l'issue de la caractérisation, nous disposons d'une description unique pour chaque structure déterminée par l'algorithme de contour actif. Nous avons utilisé 2 descripteurs, dont :

- Area : Le nombre de pixels dans la région.
- Périmètre : La longueur de la limite de l'objet.

4.3.4. Classification des machines vectorielles SVM

Nous utilisons un classifieur SVM non linéaire à deux classes (voir chapitre 3). Une étape de classification qui divise aléatoirement la base de données en un groupe d'apprentissage (d'entraînement) et un groupe de test ou de validation.

➤ Les paramètres de notre classifieur SVM sont :

- Les fonctions : « Kernel », « RBF », « BoxConstraint »
- L'erreur : $0.01 (1e^{-2})$

4.3.4.1. Évaluation des performances de classification

Dans cette partie, nous présentons les résultats obtenus lors de la phase de validation. Les matrices de confusion, également appelées tableaux de contingence, sont souvent utilisées pour évaluer les performances d'un système.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

A. La matrice de confusion :

Une matrice de confusion est un résumé des résultats de prédictions sur un problème de classification. Les résultats sont ainsi comparés avec les valeurs réelles. Ceci permet non seulement de savoir quelles sont les erreurs commises, mais surtout le type d'erreurs commises.

		Classe réelle	
		-	+
Classe prédite	-	True Negatives (vrais négatifs)	False Negatives (faux négatifs)
	+	False Positives (faux positifs)	True Positives (vrais positifs)

Tableau 4.1: Calcul d'une matrice de confusion

Vrai positif (VP) : est le nombre de patient malade qui est classé malade.

Faux négatif (FN) : est le nombre de malade qui sont classé non malade.

Faux positif (FP) : est le nombre de patient non malade qui sont classé malade.

Vrai négatif (VN) : est le nombre de patient non malade qui sont classé non malade.

Dans une matrice de confusion, VP et VN représentent les décisions correctes prises par le classifieur, alors que FN et Les FP sont des erreurs de classification.

B. Evaluation des performances

- **Précision**

Le critère de performance le plus utilisé est le taux de classification correct, connu sous le nom de précision.

Il est possible de calculer cette précision comme suit :

$$\text{Précision} = \left(\frac{VP+VN}{VP+VN+FP+FN} \right) * 100$$

- **Spécificité**

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Est la proportion de VN sur le nombre total de cas négatifs.

La spécificité peut être calculée comme suit :

$$\text{Pour notre cas : Spécificité} = \frac{VN}{(VN+FP)} * 100$$

Divisé la base de données comme suit :

1^{er} Cas :

- ✓ La base d'apprentissage est formée de 48 images (85 %) prises aléatoirement de la base de données.
- ✓ La base de validation ou de test est constituée de 8 images (15 %) prises aléatoirement de la base de données.

Les résultats obtenus sont présentés dans la matrice de confusion suivante :

	Bénigne	Maligne
Bénigne	7	0
Maligne	0	1

Tableau 4.2: matrice de confusion de notre modèle SVM .

• Précision

$$\text{Précision} = \left(\frac{VP+VN}{(VP+VN+FP+FN)} \right) * 100$$

$$\text{Précision} = \left(\frac{7+1}{8} \right) * 100 = 100\%$$

• Spécificité

$$\text{spécificité} = \frac{VN}{(VN + FP)} = \left(\frac{7}{7 + 0} \right) * 100 = 100\%$$

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Remarque :

On remarque d'après les résultats obtenus dans la matrice de confusion que le système a estimé 7 masses comme étant bénignes (VN) et une autre comme étant maligne (VP). Donc on peut dire que notre système a très bien reconnu les exemples présentés comme le montre bien la valeur de la précision obtenue (100 %).

2^{ème} Cas :

- ✓ La base d'apprentissage est formée de 40 images (70 %) prises aléatoirement de la base de données.
- ✓ La base de test (validation) est constituée de 16 images (30 %) prises aléatoirement de la base de données.

Les résultats obtenus sont présentés dans la matrice de confusion suivante :

	Bénigne	Maligne
Bénigne	13	1
Maligne	1	1

Tableau 4.3: matrice de confusion de notre modèle SVM (2^{ème} cas)

• Précision

$$\text{Précision} = \left(\frac{VP+VN}{VP+VN+FP+FN} \right) * 100$$

$$\text{Précision} = \left(\frac{13+1}{16} \right) * 100 = 87.5\%$$

• Spécificité

$$\text{spécificité} = \frac{VN}{VN + FP} = \left(\frac{13}{13 + 1} \right) * 100 = 92.86\%$$

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Remarque :

On remarque d'après les résultats obtenus dans la matrice de confusion ci-dessus que le système a estimé 13 masses comme étant bénignes (VN) et une autre comme étant maligne (VP). Le système proposé n'a pas bien reconnu tous les exemples présentés comme le montre bien la valeur de la précision obtenue (87.5 %). On remarque bien que la précision a diminué et ceci est dû au nombre d'images constituant la base d'apprentissage qui est moins important dans le 2ème cas (70 %).

4.4. Implémentation et présentation du (CAOx)

Afin de faciliter le système d'aide au diagnostic assisté par ordinateur, nous avons implémenté une interface graphique qui permet à l'utilisateur d'effectuer un prétraitement, une segmentation ou une classification. Notre (CAOx) aide le radiologue à confirmer sa décision en cas de doute par exemple. Il peut également être utilisé par des étudiants en radiologie ou des médecins inexpérimentés pour apprendre à analyser les mammographies.

4.4.1. Environnement matériel et logiciel de programmation

Notre application a été réalisée sur un ACER, 2Go de RAM sous Windows 7.

Le langage utilisé est : Matlab version 2017.

4.4.2. Présentation de l'interface graphique de notre projet :

L'interface est composée d'une fenêtre montrant le menu principal, constitué de des boutons permettant d'accéder aux différentes fonctionnalités telles que :

- L'acquisition de l'image et la détermination de la région d'intérêt (manuellement par la souris)
- L'élimination du bruit et amélioration du contraste
- La segmentation par contour actif
- La classification par SVM

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)



Figure 4.11: menu de l'application

En cliquant, sur le bouton Importer l'image nous avons choisi l'image mdb015, ensuite en cliquant sur le bouton «Élimination de bruit et amélioration du contraste» l'interface suivante apparait (figure 4.12) montrant des images prétraitées de l'image chargée (originelle) :

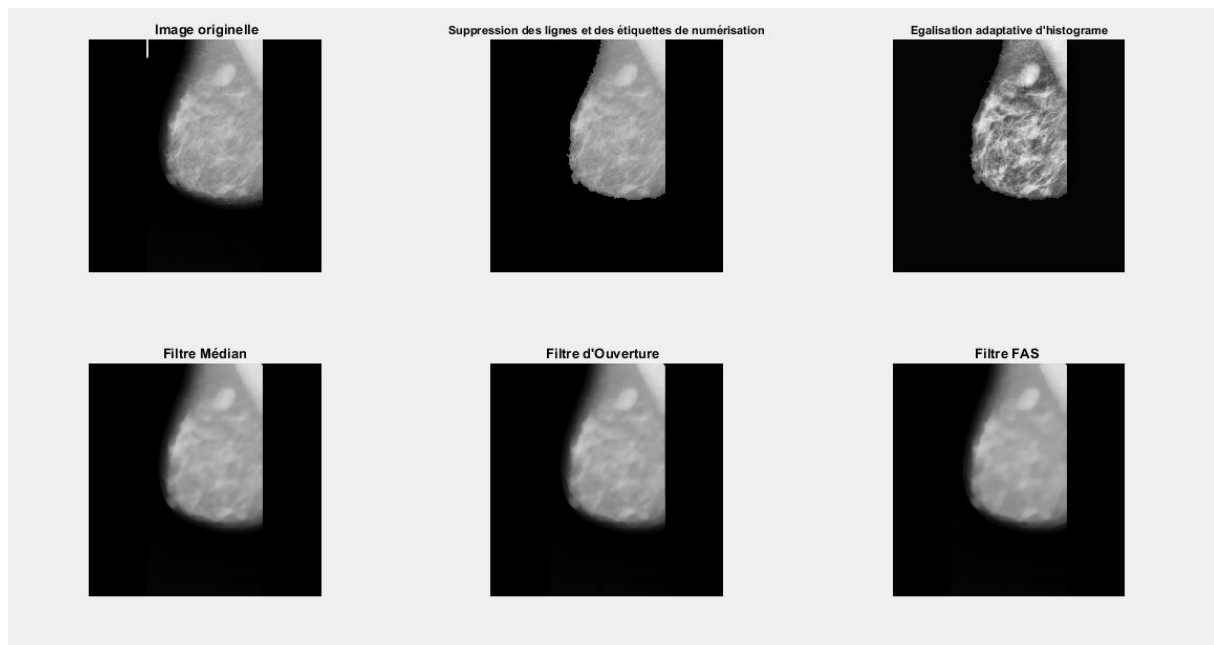


Figure 4.12 : résultat de l'élimination de bruit et amélioration du contraste de l'image mdb015.

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Nous avons refait le même travail sur l'image mdb202, l'interface suivante apparaît (figure 4.13) montrant des images prétraitées de l'image chargée (originelle) :

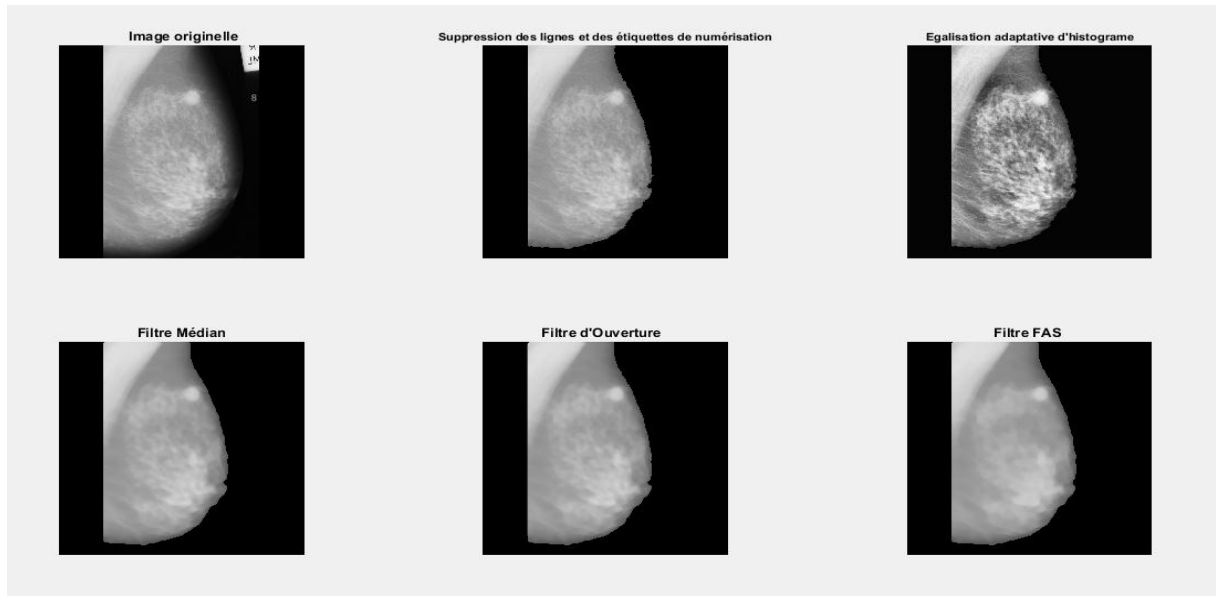


Figure 4.13 : résultat de l'élimination de bruit et amélioration du contraste de l'image mdb202

En cliquant sur le bouton « Contour actif » l'interface suivante apparaît (figure 4.15) faisant apparaître la segmentation par contour actif de l'image mdb015

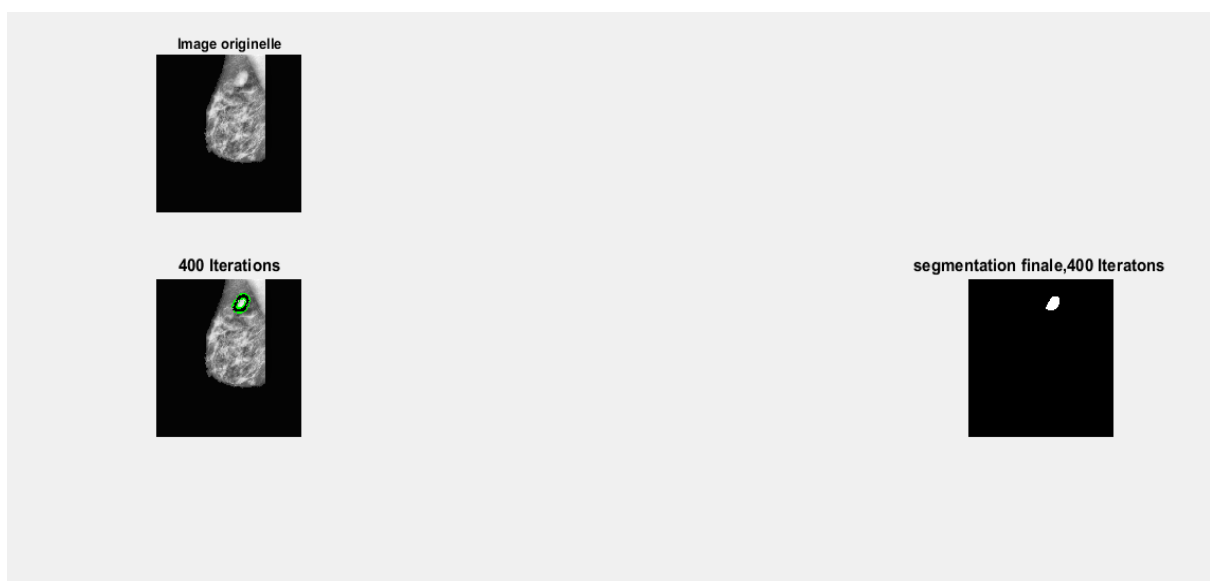


Figure 4.14: résultat de contour actif de l'image mdb015

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)

Nous avons refait la même chose mais avec d'autre image donc l'interface suivante apparait (figure 4.15) faisant apparaitre la segmentation par contour actif de l'image mdb202 :

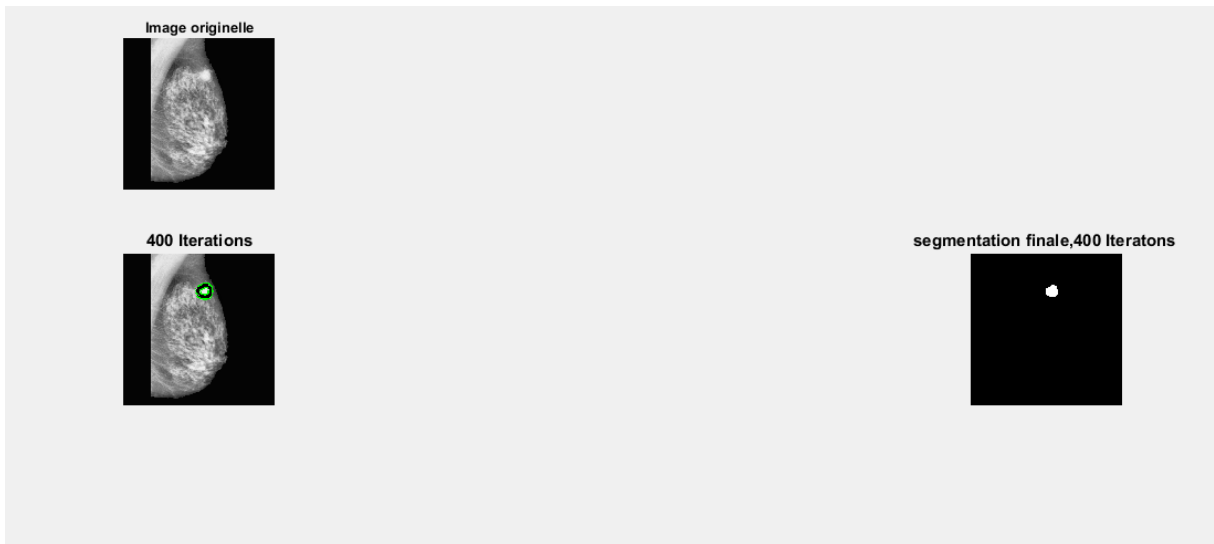


Figure 4.15 : résultat de contour actif de l'image mdb202.

En cliquant sur le bouton « Classification » l'interface suivante apparait (figure 4.16) affichant le résultat de la classification comme étant bénigne ou maligne :



Figure 4.16: résultat de la classification de l'image mdb015

Chapitre 04 Conception et implémentation d'un système d'aide au diagnostic assisté par ordinateur (CAOx)



Figure 4.17: résultat de la classification de l'image mdb202

4.5. Conclusion

Dans ce chapitre, nous avons mis au point un système d'aide au diagnostic assisté par ordinateur basé sur l'utilisation du contour actif et les machines SVM. L'algorithme proposé donne d'excellents résultats. En effet, afin d'évaluer ses performances, nous avons considéré deux cas de figures de la base de données et les résultats ont montré qu'il faut au moins prendre 80 % de la base de données comme étant la base d'apprentissage afin d'avoir un taux de reconnaissance très élevé.

Conclusion

Générale

Conclusion générale et perspective

Dans ce mémoire, nous nous intéressons à l'étude des systèmes semi-automatisés de diagnostic assisté par ordinateur pour le diagnostic du cancer du sein, en particulier la description et la classification des masses mammaires. Les chercheurs en imagerie se concentrent sur différentes étapes de la chaîne de diagnostic assisté par ordinateur pour améliorer les résultats de la classification des masses comme étant malignes/bénignes.

Dans ce cadre, des recherches ont été menées sur l'utilité des outils et du prétraitement des images mammographiques. En effet, la partie de prétraitement est basée sur un ensemble de techniques (filtrage, opérations morphologiques) pour nettoyer l'image afin de la préparer à la prochaine phase du processus de reconnaissance de forme.

Par la suite, nous avons démontré la fiabilité de la segmentation par contour actif en élaborant un algorithme pour la détection des masses dans la glande mammaire et faciliter l'opération de classification.

En dernier lieu, afin d'identifier la tumeur selon qu'elle soit bénigne ou maligne, nous avons mis au point un classifieur basé sur l'utilisation des machines SVM (Support Vector Machines). Les résultats obtenus sont très intéressants, même s'ils sont loin des performances des meilleures méthodes proposées dans la littérature.

Ce mémoire nous permet de dégager plusieurs perspectives que nous résumons comme suit :

1. Les propriétés morphologiques que nous utilisons sont encore très simples, elles peuvent donc être enrichies d'autres propriétés de niveau supérieur : texture, luminosité, attributs...

2. La méthode de la machine à vecteurs de support (SVM) sélectionnée ne permet de classer que deux classes, une troisième classe peut être ajoutée, qui présente des seins normaux ne contenant pas de tumeurs. Il s'agit d'un type de machine SVM dit "multi-class".

Conclusion générale et perspective

3. Nous pouvons tester notre système sur une autre base de données plus puissante telle que DDSM.

4. Nous pouvons également utiliser les réseaux de neurones comme classifieur pour une étude comparative avec des machines SVM.

La morphologie mathématique

A.1. Introduction

La morphologie mathématique est une science de la forme et de la structure. Le principe de base de la morphologie mathématique est de comparer les objets que l'on veut analyser à un autre objet de référence, de taille et de forme connues, appelé élément structurant. En quelque sorte, chaque élément structurant fait apparaître l'objet sous une nouvelle forme.

Fondamentalement, la morphologie mathématique a un caractère ensembliste et ceci correspond à la démarche dont nous disposons pour analyser quantitativement une image. [47] [48]

Avant de définir les différents opérateurs de la morphologie mathématique, nous définissons d'abord l'élément structurant qui est un élément de base de toute opération de la morphologie mathématique.

➤ L'élément structurant

L'élément structurant (ES) est un masque de forme quelconque dont ses éléments forment un motif, il est défini comme suit :

Soit B un sous-ensemble de E , nommé élément structurant. Si x est un élément de E , alors on définit l'ensemble Bx , le déplacement de B en chaque point x de l'espace E :

$$Bx = \{b + x \mid b \in B\}$$

On introduit aussi le symétrique de B , noté $BS : Bs = \{-b, \forall b \in B\}$ si l'élément structurant est symétrique, on a : $BS = B$.

Annexe

A.2. Les opérateurs morphologiques

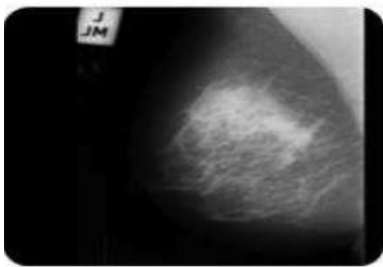
➤ Erosion

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 se pose la question suivante : est-ce que B est complètement inclus dans X ?

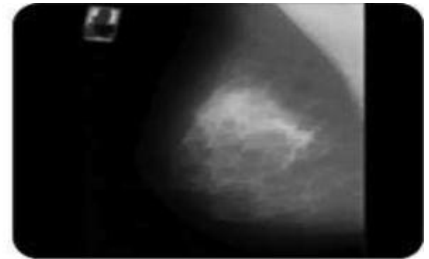
L'ensemble des positions x correspondant à une réponse positive forme le nouvel ensemble Y , appelé érodé de X par B , noté $\varepsilon_B(X)$ (figure A.1).

Cet ensemble satisfait l'équation :

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



a) Image mammographique initiale



b) Image mammographique érodée

Figure A.1: Erosion d'une image mammographique.

Après une érosion et selon la figure ci-dessus nous pourrions dire que :

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.

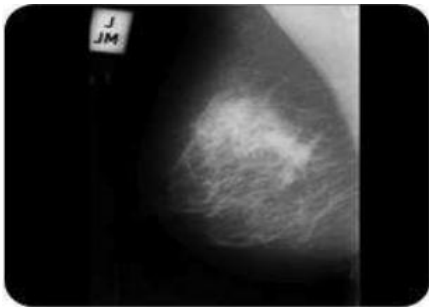
Annexe

➤ Dilatation

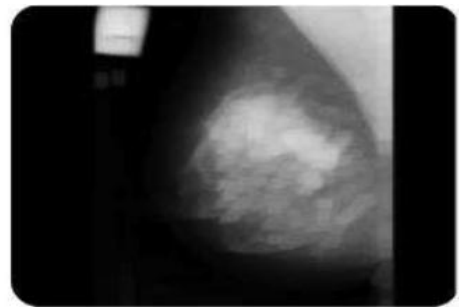
L'opération de dilatation se définit de manière analogue à l'érosion. En prenant le même élément structurant B , on se pose pour chaque point x la question : " Bx touche-t-il l'ensemble X ? ", c'est à dire, y a-t-il une intersection non vide entre Bx et X ?

L'ensemble des points de l'image (figure A.2) correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X , noté comme suit :

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



a) Image mammographique initiale



b) Image mammographique dilatée

Figure A.2: Dilatation d'une image mammographie.

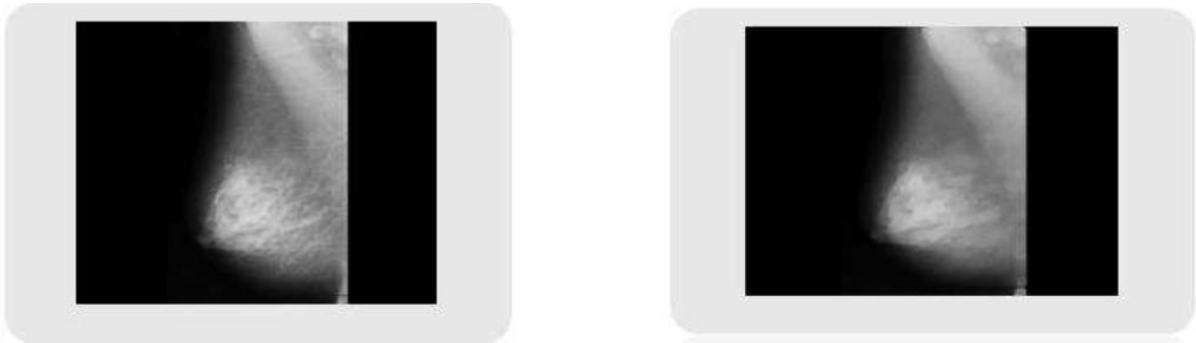
➤ Ouverture

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 :

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

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 (voir figure A.3). La transformation par ouverture adoucit les contours, coupe les isthmes étroits, supprime les petites îles et les caps étroits.

Annexe



a) Image mammographique initiale b) Image mammographique avec ouverture

Figure A.3: Ouverture d'une image mammographie par un élément structurant circulaire.

L'opération d'ouverture 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, de ne pas conserver nécessairement la topologie.

➤ 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 :

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

Un ensemble ferme est également moins riche en détails que l'ensemble initial (figure A.4). La transformation par fermeture bouche les canaux étroits, supprime les petits lacs et les golfes étroits.

Annexe



a) Image mammographique initiale

b) Image mammographique avec fermeture

Figure A.4 : 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 souvent la taille et la forme.
- Ne pas conserver nécessairement la topologie.
- Souder les formes proches en particulier.

➤ Filtres Alternés Séquentiels (FAS)

A partir des filtres morphologiques élémentaires, tels que les ouvertures et les fermetures, on définit des filtres morphologiques plus évolués. Parmi ceux-ci, les Filtres Alternés Séquentiels (FAS) qui sont sans doute les plus populaires. Ils permettent de réduire la complexité d'une image sans dégrader les contrastes les plus importants. On définit le Filtre Alterné Séquentiel Noir de taille n , noté $FAS_N(n)$, comme une itération de succession d'ouverture et de fermeture de taille croissantes. [49]

$$FAS_N(n) = F_n O_n \dots F_2 O_2 F_1 O_1$$

Où O_i et F_i sont une famille d'ouvertures et une famille de fermetures par un élément structurant de même forme et de taille croissante. On peut changer également l'ordre des ouvertures et des fermetures pour obtenir un Filtre Alterné Séquentiel Blanc (FASB) :

$$FAS_B(n) = O_n F_n \dots O_2 F_2 O_1 F_1$$

Annexe

Ces filtres FAS qui sont gourmands en temps de calcul [50] conduisent à une bonne réduction du bruit grâce à une élimination progressive des pics et des creux de faible surface. En pratique, le dernier élément structurant utilisé n'est pas choisi librement. Il est déterminé en fonction de la taille minimale des objets de l'image que l'on veut conserver après filtrage.

L'exemple suivant (figure A.5) utilise le Filtre Alterné Séquentiel (FAS) noir et blanc sur une image maligne [50], avec une variation des paramètres sur les images mammographiques.

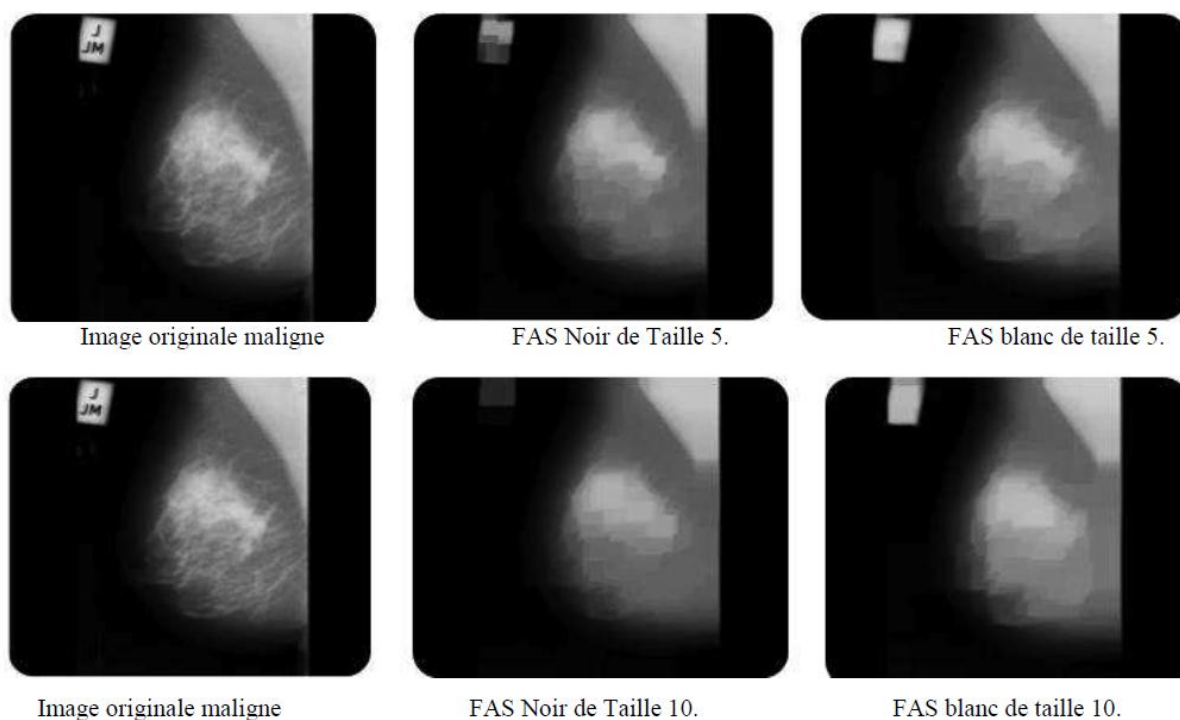


Figure A.5: Application du Filtre Alterné Séquentiel noir et blanc sur une image mammographique maligne.

Bibliographie

[01] A.Lecorgne : « Le rôle de pharmacien d'officine dans la prise en charge du cancer du sein après chirurgie mammaire ». Thèse de doctorat, Université de Bourgogne UFR des sciences de santé circonscription pharmacie, 25-11-2016.

[02] I.Hadjij : « Approche morphologique pour la segmentation d'images médicales, application à la détection des lésions mammaires ». Mémoire de magister en électronique biomédicale, Université Aboubekrbelkaid-Telemcen, Faculté des technologies, Département de génie électrique et électronique, 06-07-2011.

[03] W.Eziddin : « Segmentation itérative d'images par propagation de connaissance dans le domaine possibiliste : application à la détection de tumeurs en imagerie mammographique ». Thèse de doctorat, Université européenne de Bretagne, département images et traitement de l'information, N°2012telb0231, 14-04-2012.

[04] Z.Tahri : « Etude histopathologie et immunohistochimie des cancers mammaires : à propos de 50 cas ». Mémoire de magister, option cancer et environnement, Université d'Oran ES-SENIA, faculté des sciences, département de biologie, 2007-2008.

[05] M. Guillaume : « Anatomie ». Cours, CHU Pontchaillou et Université de Rennes 1, Rennes, 2010.

[06] <http://www.nuclearord.fr/informations-reperage-scintigraphique-du-ganglion-sentinelle-4.html>

[07] A.S.Collins : « Comprendre l'anatomie du sein pour comprendre le cancer du sein ». BIRON B groupe santé, Centre du savoir de l'information médicale, Les petits guides bien-être et santé Biron, 2015.

[08] J.Zucman-rossi et J.C.Nault : « Comment une cellule devient cancéreuse et qu'on peut faire pour l'éviter ». The conversation (academic rigor, journalistic flair), 20-12-2016 mise à jour 02-02-2017.

[09] P.Fumleau et D.Mignard sous la coordination de M.Lanta : « Les cancers du sein ». Ligue contre le cancer, Ref : BRK006, 12-2017.

[10] BI-RADS. Breast Imaging reporting and Data System. Atlas d'imagerie du sein Mammographie. Deuxième édition française basée sur la quatrième édition américaine.

[11] BI-RADS. Breast Imaging reporting and Data System. Atlas d'imagerie du sein Mammographie. Deuxième édition française basée sur la quatrième édition américaine.

[12] [En ligne] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>.

[13] [En ligne] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>.

[14] S.Chabane et N.Haddache : « Segmentation d'image mammographies ». Mémoire de master, Université Abderrahmane Mira Bejaia, faculté des sciences exactes, Département informatique, 2016-2017.

[15] M.P.Sampat, M.Markey, A.C.Bovik : « Détection et diagnostic assisté par ordinateur en mammographie ». Manuel de traitement d'image et de vidéos, volume 2, pages 1195-1217.

[16] J. Zander and J. Baltzer, S. Verlag, Berlin: «Morphologic analysis of microcalcifications». Early Breast Cancer, Lanyi, M, 1985.

[17] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., and Sickles., E. A : «American college of radiology breast imaging reporting and data system». Troisième édition française, réalisée par SFR (Société Française de Radiologie), 2003.

[18] Med Phys : « L'imagerie du sein de rapports et système de données (BI-RADS) descripteurs de composition du sein : le développement de mesure automatisé pour mammographie numérique plein champ ». Medical physics, 40 (11): 113502, 11-2013.

[19] Ikeda DM, Hylton NM, Kinkel K et al. Development, standardization and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance (MR) imaging studies. J Magn Reson Imaging 2001 ;13 :896-902.

[20] Bailar, 1976] Bailar, J. C. (1976). Mammography : a contrary view. Annals of Internal Medicine, 84(1) :77-84.

[21] Ikeda DM, Hylton N, Kinkel K et al. Development, standardization and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance (MR) imaging studies. 85th Annual Meeting of the Radiological Society of North America (RSNA), Chicago, USA, December 1999

[22] J.Radiol : « Analyse radiologique des cancers d'intervalle connus, après deux ans de campagne de dépistage de masse organisé (DMO) du cancer du sein en Ile-et-Vilaine ». Edition française de radiologie, Paris, vol : 79, pages : 1379-1386, 1998.

[23] A.Tradivon : « Quels risques, pour quels femmes ?, Densité mammaire et cancer du sein ». Conférence 30es journées de la SFSPM, LA BAULE, 11-2008.

[24] [Allen et al., 2003] Allen, B. H., Oxley, M. E., and Collins, M. J. (2003). A universal segmentation platform for computer-aided detection. In Digital Mammography, pages 164-168. Springer.

[25] [Baker et al., 2003] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., and Soo, M. S. (2003). Computer-aided detection (cad) in screening

mammography : sensitivity of commercial cad systems for detecting architectural distortion.

[26] [En ligne] <http://peipa.essex.ac.uk/info/mias.html>.

[27] D’Osri, C., Bassett, L., Berg, W. et al. (2003). Breast imaging reporting and data system : Acr bi-rads, Technical report, American College of Radiology.

[28] Monsees BS. The Mammography Quality Standards Act : an overview of the regulations and guidance. *Radiol Clin North Am* 2000 ; 38 :759–772 [Crossref] [Medline] [Google Scholar].

[29] H. Hwang, B. Choi, H. Lee, H. Yoon, Y. Nam, and H. Choi, Multiresolution wavelettrans formed image analysis of histological sections of breast carcinomas, *Cellular Oncology*, vol. 27, 2005.

[30] [Baker et al., 2003] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., and Soo, M. S. (2003). Computer-aided detection (cad) in screening mammography :sensitivity of commercial cad systems for detecting architectural distortion. *AmericanJournal of Roentgenology*, 181(4) :1083–1088.

[31] Benyahia Souad Née Belaidi, “Application de la classe des méthodes d’apprentissagestatistique SVM (support vector machine) pour la reconnaissance des formes dans lesimages”, mémoire pour obtenir le titre de magister en informatique 2012.

[32] www.ai.univ-paris8.fr/amsi/PGSI1516S1/docs/wiki_Filtre_median.pdf/

[33] R. D. Yapa et Koichi Harada, «Breast Skin-Line Estimation and Breast Segmentationin Mammograms using Fast marching Method,» *International Journal of Biologicaland Life Sciences*, vol. 3, %11, pp. 54-62, 2007.

[34] B.E. Boser, I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[35] I. Guyon, B. Boser, and V. Vapnik. Automatic capacity tuning of very large vc-dimension classifiers. In *Advances in neural information processing systems*. Citeseer, 1993.

[36] V.N. Vapnik. *Statistical Learning Theory*. Edition Wiley, 1998.

[37] V.N. Vapnik. *The nature of statistical learning theory*. Springer, 2000.

- [38] Mohamadally Hasan, Fomani Boris : " SVM machine a vecteurs de support ou separateur a vaste marge ". BD Web, ISTY3, Versailles St Quentin, France, janvier 2006.
- [39] W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Dept. of Mathematics, Univ. of Chicago, 1939.
- [40] H.W. Kuhn and A.W. Tucker. Nonlinear programming. In University of California Press, editor, 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics, pages 481–492, 1951.
- [41] B. Scholkopf and A.J. Smola. Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2002.
- [42] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 209 :415–446, 1909.
- [43] H. König. Eigenvalue distribution of compact operators. Birkhäuser Verlag Basel Boston Stuttgart, 1986.
- [44] R. Herbrich. Learning Kernel Classifiers : Theory and Algorithms. MIT Press, 2002.
- [45] K.P Soman and R. Loganathan and V. Ajay. Machine learning with SVM and other kernel methods. Eastern Economy Edition, 2009.
- [46] Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2) : 2009.
- [47] S. Beucher, "Segmentation d'images et morphologie mathématique", thèse de doctorat, école Nat. Sup. des Mines de Paris, 1990.
- [48] M. Coster et J.L. Cherman, "Traitement d'image par morphologie mathématique Application au domaine médical et industriel", Thèse de doctorat, université Blaise Pascal, 1987.
- [49] HAOUZI Badra, "Extraction de réseaux linéiques à partir des images à haute résolution", Mémoire en vue de l'obtention du diplôme de magister en Télécommunications et Informatique Spatiales 2012.
- [50] The mini-MIAS database of mammograms, available online en 2014 at: <http://peipa.essex.ac.uk/info/mias>