



وزارة التعليم العالي والبحث العلمي  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE  
LA RECHERCHE SCIENTIFIQUE  
جامعة عبد الحميد ابن باديس مستغانم  
Université Abdelhamid Ibn Badis Mostaganem  
كلية العلوم والتكنولوجيا  
Faculté des Sciences et de la Technologie  
DEPARTEMENT DE GENIE ELECTRIQUE



N° d'ordre : M 17/10/2023

## MEMOIRE DE FIN D'ETUDES

Présenté pour obtenir le diplôme de MASTER ACADEMIQUE

**Filière :** Électronique

**Spécialité :** Électronique des Systèmes Embarqués

Présenté par :

**GOUCEM ASMA**

# La détection et classification des masses mammaires par SVM

Soutenu le 17 / 10 / 2023 devant le jury composé de :

President	M. Mimi	Prof	Université de Mostaganem
Examineur:	M. Bentoumi	M.C.A	Université de Mostaganem
Encadreur:	K. Berradja	M.C.A	Université de Mostaganem
Co-encadreur :	A. Mehidi	M.C.B	Université de Mostaganem

Année Universitaire 2022/2023

# Dédicaces

*Tout d'abord, je tiens à remercier DIEU De m'avoir donné la force et le courage de mener à bien ce modeste travail.*

*A mes chers parents pour avoir été à mes côtés tout au long de mon parcours académique et pour m'avoir encouragé dans ce long cheminement.*

*A mes frères et surtout mon frère Elias, qui m'a accompagné tout au long de mes études et m'a soutenu.*

*A toute ma famille*

*A tous mes professeurs et amis qui ont contribué à la réalisation de cet humble travail*

*Asma Asia Goucem*

# REMERCIEMENTS

En premier lieu, je suis reconnaissante à Dieu Tout-Puissant d'avoir été mon guide, de m'avoir éclairé par la connaissance et de m'avoir donné la volonté, la force et le courage de faire cet humble travail.

J'exprime ma gratitude à Dr. K. BERRADJA et Dr. A. MEHIDI, pour m'avoir guidé tout au long de ce travail. Je les remercie pour leur grand professionnalisme, leurs conseils, leur soutien et leurs disponibilités continus pour accomplir ce travail.

Je remercie Professeur M. Mimi, pour avoir accepté de rendre compte de ce travail et de nous honorer de la présidence du jury.

Je tiens également à remercier Dr. M. Bentoumi, pour avoir accepté d'examiner ce mémoire.

Enfin, je tiens à remercier ma famille pour son soutien continu et ses encouragements durant la préparation de ce mémoire, et plus particulièrement mes parents pour leur compréhension, leur présence et leur patience. Merci d'avoir fait de moi ce que je suis aujourd'hui. à tous mes amis de l'université et d'ailleurs, à tous ceux qui ont contribué, de près ou de loin, à l'accomplissement de cet humble travail.

## RESUME

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 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 précoces 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 prétraitement qui vise à réduire le bruit et améliorer le contraste. Ensuite, la segmentation dans laquelle nous utilisons la méthode k-means pour isoler les structures suspectes. Enfin, l'étape de classification par les machines à vecteur support (SVM) est effectuée, permettant d'identifier le type de l'anomalie observée comme étant maligne ou bénigne. Notre système a utilisé la base de données Mini MIAS.

**Mots clés** : cancer du sein, segmentation par k-means, systèmes d'aide au diagnostic, classification par SVM.

## ABSTRACT

Breast cancer is one of the most common causes of death in women. One in eight women is at risk of developing the disease in their lifetime. It is essential to detect the disease early, before it has a chance to spread. Due to his late diagnosis, he often results in expensive and expensive treatment. Thanks to digital mammography and diagnostic support systems (CAD), the chances of survival are greatly increased. The challenge is to detect cancers that radiologists may miss. This makes it the main research directorate for our project. In this work, we are developing a system for detecting and classifying suspicious anomalies on mammograms. The proposed system is divided into several stages: a pre-treatment phase aimed at reducing noise and improving contrast. Next, the segmentation in which we introduce the k-means method to isolate suspicious structures. Finally, the classification step, in which an SVM classification is performed, allowing to identify the type of the observed anomaly (malignant or benign). Our system has been tested on images from the Mini MIAS database.

**Keywords**: breast cancer, k-means segmentation, SVM, diagnostic support systems.

# TABLE DES MATIERES

<b>INTRODUCTION GENERAL.....</b>	<b>1</b>
<b>CHAPITRE I : .....</b>	<b>4</b>
I .1 Introduction.....	5
I .2 Anatomie du sein chez la femme .....	5
I .3 Le cancer du sein.....	6
I .4 Dépistage du cancer du sein.....	7
I .4.1 La mammographie :.....	7
I .5 Types d’anomalies dans le sein.....	9
I .5.1 Les calcifications : .....	9
I.5.2 Les masses : .....	9
I .6 Diagnostic du cancer du sein assisté par ordinateur (DAOx) .....	12
I .6.1 L’importance des systèmes de diagnostic par ordinateur :.....	13
I .7 Les bases d’images de mammographie : .....	13
I .7.1 Mammographie Image Analysais Society (MIAS) .....	13
I .7.2 Digital Database For Screening Mammography (DDSM).....	14
I .7.3 Banco Web LAPIMO .....	14
I.8 Conclusion.....	15
<b>CHAPITRE II : Traitement des Images de Mammographies .....</b>	<b>16</b>
II.1 Introduction.....	17
II .2 Traitement d’image mammographique.....	17
II .2.1 Prétraitement des images mammographies .....	17
II.3 Segmentation des images mammographie .....	22
II.3.1 Algorithme de k-Means.....	23
II.4 Conclusion : .....	25
<b>CHAPITRE III : .....</b>	<b>26</b>
<b>Les machines à Vecteur Support .....</b>	<b>26</b>
<b>( SVM).....</b>	<b>26</b>

## Table des Matières

---

III.1 Introduction.....	27
III.2 Support Victor machine (SVM) .....	27
III.3 Principe de fonctionnement .....	28
III.4 Formalisme des SVM .....	30
III.5 Entraînement des SVM .....	31
III.5.1 Calcul de la marge .....	31
III.5.2 Maximisation de la marge.....	32
III.6 SVM non linéaire et la fonction noyau.....	33
III.6.1 Exemple de séparation non linéaire .....	34
I.7 Conclusion .....	35
<b>CHAPITRE IV : Conception et Implémentation d'un Système d'aide au Diagnostic assisté par Ordinateur (CAOx) .....</b>	<b>36</b>
IV.1. Problématique .....	37
IV.2. Base de données .....	37
IV.3 Conception du système d'aide au diagnostic assisté par ordinateur.....	37
IV.3. 1.Prétraitement des données .....	40
IV.3. 2. Segmentation .....	44
IV.3. 3. Descripteurs de forme.....	46
IV.3.4. Classifieur SVM.....	46
IV.4. Résultats obtenus .....	48
IV.5. Conclusion .....	50
<b>CONCLUSION GENERAL .....</b>	<b>51</b>
<b>BIBLIOGRAPHIE.....</b>	<b>53</b>

## LISTE DES FIGURES

**Figure I.1 :** L'anatomie du sein chez la femme.

**Figure I.2 :** Exemple de mammographie.

**Figure I.3 :** Visualisation lever du bras pour oblique médio latérale.

**Figure I.4 :** Visualisation craniocauda (cc).

**Figure I.5 :** Les différentes formes possibles d'une masse.

**Figure I.6 :** Les types de densité mammaire.

**Figure I.7 :** Les différents contours d'une masse mammaire.

**Figure II.1 :** Traitement d'une image mammographique.

**Figure II.2 :** Exemple de suppression des artefacts d'une mammographie.

**Figure II.3 :** Élimination de bruit avec le filtre médian.

**Figure II.4 :** Exemple d'image améliorée par égalisation d'histogramme adaptative.

**Figure II.5 :** Ouverture classique d'une image mammographique.

**Figure II.6 :** Application de filtre alterné séquentiel sur une image mammographie.

**Figure II.7 :** Organigramme de méthode de k-means.

**Figure II.8 :** Exemple de segmentation de masse par k-means.

**Figure III.1 :** Le séparateur idéal correspond intuitivement à l'hyperplan qui passe au milieu c'est le séparateur de marge maximale.

**Figure III.2 –** Les vecteurs de support se trouvent à une distance égale a la marge d'un côté ou de l'autre de l'hyperplan de séparation.

**Figure III.3 –** Astuce à noyaux : projeter les données dans un espace de dimension beau- coup plus grande, où elles deviennent séparables linéairement.

**Figure III.4 –** Exemple de changement de l'espace de données.

**Figure IV.1 :** La séparation idéale correspond intuitivement à l'hyperplan qui passe est milieu, c'est la séparation de marge maximale.

**Figure IV.2:** Les vecteurs de support se trouvent à une distance égale à la marge d'un côté ou de l'autre de l'hyperplan de séparation.

**Figure IV.3 :** Suppression des artefacts : (a) et (b) images originelles, (c) et (d) après élimination des étiquettes.

**Figure IV.5 :** Suppression du bruit : (a) images obtenues précédemment, (c) images filtrées.

**Figure IV.6 :** Amélioration du contraste. (a) les images filtrées, (c) rehaussement de contraste par une ouverture.

**Figure IV.7 :** Amélioration du contraste : (a) et (b) images filtrées par ouverture, (c) et (d) images avec rehaussement de contraste par filtre FAS.

**Figure IV.8 :** (a) et (b) sont des images segmentées par K-Means, (c) et (d) sélection de la zone d'intérêt (ROI) et (e) et (f) extraction des ROI.



# **INTRODUCTION GENERAL**

## INTRODUCTION GENERALE

Le cancer du sein est le cancer le plus fréquent chez les femmes avec 450000 décès chaque année dans le monde. En Algérie, il représente près de 50 % des cancers gynécologiques. Le diagnostic tardif entraîne souvent un traitement onéreux et coûteux, avec un taux de mortalité élevé. Cependant, seul le traitement dès le premier stade de la tumeur peut réduire la mortalité et augmenter les chances de guérison de ce type de cancer. Pour garantir 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. Par exemple, effectuer un dépistage systématique tous les deux ans s'est avéré être une méthode très efficace et réduire le taux de mortalité d'environ 35 %. En raison d'une ou deux mammographies par sein, deux ou quatre mammographies sont effectuées par patiente, ce qui entraîne une augmentation exponentielle du nombre de mammographies effectuées, rendant la tâche d'interprétation des radiologues plus difficile. Le taux de détection du cancer du sein peut être augmenté d'environ 15 % après une seconde lecture. Avec l'augmentation du nombre de mammographies effectuées au cours des dernières décennies, divers efforts de recherche sont déployés pour détecter automatiquement les lésions (micro calcifications ou masses) ou pour fournir un deuxième avis sur la lésion détectée grâce à des systèmes de diagnostic assistés par ordinateur. Cela a conduit à la création de diverses techniques d'imagerie mammographique, toutes sont basées sur la classification BI-RADS (Breast Imaging Reporting and Data System) de l'American College of Radiology (ACR) qui classe les masses en deux catégories, bénignes et malignes.

À cet effet, notre objectif est de développer un système d'aide semi-automatique basé sur :

- L'utilisation d'une méthode de segmentation et de détection semi-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.

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

Le premier chapitre est dédié à la définition des concepts fondamentaux de l'anatomie du sein et à la description de ses principaux composants. De plus, ce chapitre contient un aperçu sur les masses mammaires, leurs caractéristiques et de leurs classifications.

Le deuxième chapitre est consacré à l'analyse des images de mammographie. Les méthodes de prétraitement standard du traitement d'image, telles que la réduction du bruit et l'amélioration du contraste, sont décrites en premier lieu. La phase de segmentation utilisant la méthode de k-means pour la détection et l'extraction des masses est mentionnée en deuxième lieu.

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 présenté pour détecter les masses dans les images mammographiques et les classer comme étant bénignes ou malignes. On y trouve également les résultats obtenus.

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

**CHAPITRE I :**

**Généralités sur**

**Le Cancer du Sein**

## I.1 Introduction

Le cancer du sein est le type de cancer le plus répandu chez les femmes. Il touche généralement les femmes de plus de 50 ans. Aujourd'hui, les femmes de tous âges peuvent avoir un cancer du sein. Dans ce chapitre, nous allons décrire, en premier lieu, les principales structures anatomiques du sein. Nous expliquerons ensuite le processus de dépistage du cancer du sein, puis nous passerons aux principes généraux de l'acquisition de l'imagerie mammaire. Cette introduction permettra de comprendre l'articulation des différentes étapes qui composent notre système de diagnostic proposé.

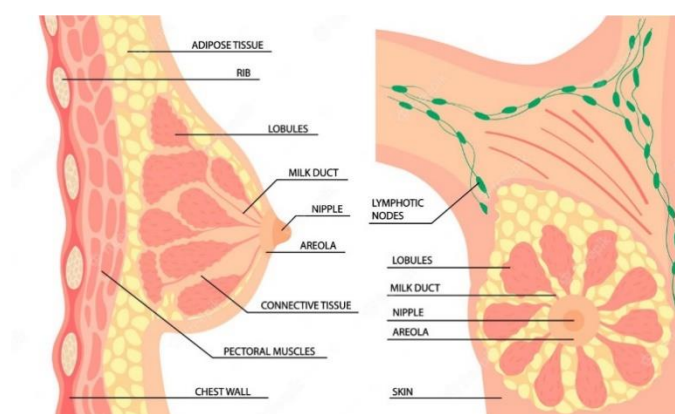
## I.2 Anatomie du sein chez la femme

Les seins sont des glandes exocrines productrices de lait entourant les glandes mammaires qui se développent continuellement tout au long de la vie d'une femme (voir le Figure I.1).

Le composant actif de la glande mammaire est constitué des vésicules mammaires (ou acini), qui se réunissent pour former un lobe [1].

Les parois alvéolaires sont composées de cellules épithéliales sécrétoires et de cellules myoépithéliales contractiles [2].

- 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 [3].
- 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 [04].
- Le mamelon est la zone au centre de l'aréole avec du lait sortant d'une extrémité [5].
- L'aréole est la surface ronde, rose ou brune qui entoure le mamelon. Il contient de petites glandes qui libèrent ou sécrètent une substance huileuse qui agit comme un lubrifiant pour le mamelon et l'aréole [6].
- Les seins sont tapissés de vaisseaux sanguins et de vaisseaux lymphatiques. Les ganglions lymphatiques de drainage mammaire sont situés dans les aisselles, les glandes supra claviculaires, sous-clavières et Mammaires internes (à l'intérieur de la cavité thoracique) [7].



**Figure I.1** – L'anatomie du sein chez la femme [08].

### **I.3 Le 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 [09].

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) [10]. 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 sans potentiel métastatique. Même si ces tumeurs ne sont généralement pas agressives envers les tissus environnants, elles peuvent parfois continuer à se développer, en appuyant sur les organes et en provoquant des douleurs ou d'autres problèmes. Dans ces situations, la tumeur est retirée, permettant à la douleur ou aux complications de s'apaiser.

#### **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.

## I.4 Dépistage du cancer du sein

Le dépistage du cancer du sein consiste à réaliser des examens de surveillance d'une personne supposée en bonne santé afin d'arriver à détecter des anomalies pouvant être des signes avant-coureurs d'un cancer, bien avant l'apparition des premiers symptômes. Cette démarche permet de réduire significativement le taux de mortalité lié au cancer de sein en améliorant les chances de guérison. Il existe différentes modalités d'imagerie permettant de détecter des anomalies mammaires, parmi lesquelles on cite.

### I.4.1 La mammographie :

Le but de la mammographie est de détecter précocement les anomalies avant même qu'elles ne provoquent des symptômes cliniques. Par conséquent, elle permet 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 la mammographie [11] (Figure I.2).



**Figure I.2 :** Exemple de mammographie [12].

Une mammographie ne vous donne pas toujours un diagnostic immédiat et définitif : elle vous

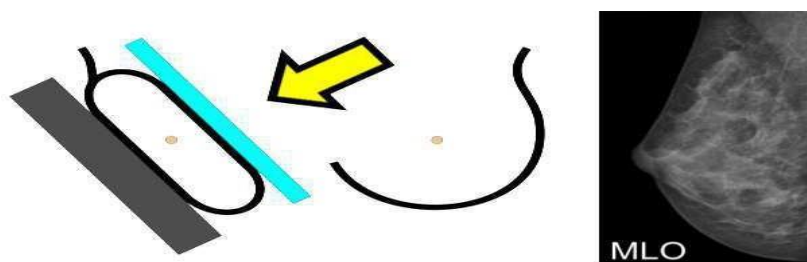
permet de voir s'il y a une anomalie dans le sein, mais elle ne vous permet pas de savoir avec certitude s'il s'agit d'un cancer.

La douleur associée à la compression du sein entre les deux plaques lors de l'examen est fréquente mais moins intense. Le premier risque après mammographie de dépistage est le sur diagnostic. Entre 20 % et 49 % des femmes qui passent régulièrement des mammographies auront au moins un résultat faussement positif après 10 examens.

La mammographie numérique est une image électronique qui est produite avant d'être enregistrée sur un ordinateur où le contraste des zones les plus denses est traité, notamment avant visualisation. La mammographie numérique a 2 positions de Visualisation:

- **Lever du bras pour Oblique Médio Latérale (MLO) :**

Lors de la visualisation du MLO, la patiente est également amenée à se pencher vers l'équipement pour une visualisation optimale des tissus. Les bras sont tournés à 45 degrés pour révéler la quantité maximale de tissu mammaire et de muscles pectoraux. Parfois, les angles sont individualisés en fonction de la taille de la poitrine et varient jusqu'à 10 degrés. L'autre sein du patient, non-imageur, est doucement pressé contre le corps et maintenu à l'écart [13] (Figure I.3).

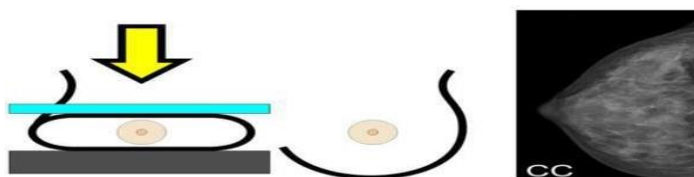


**Figure I.3 :** Visualisation lever du bras pour oblique médio latérale (MLO) [14].

- **Craniocaudal (CC) :**



Pendant la vision CC, la patiente est positionnée de sorte que le mamelon soit approximativement centré sur le 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 de sorte que les tissus inférieurs plus profonds soient inclus dans le champ de vision, et le côté imagé est poussé vers le bas pour libérer les muscles pectoraux, incluant ainsi le tissu mammaire dans le quadrant externe. La visualisation des muscles pectoraux dans la vue CC signifie qu'aucun tissu de la paroi thoracique n'a été exclu) [15] (Fig. I.4).



**Figure I.4 :** Visualisation craniocaudal (CC) [16].

## **I.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 et les masses.

### **I.5.1 Les calcifications :**

Il existe deux types de calcifications : les macro-calcifications et les micro-calcifications. Elles se traduisent, dans les clichés mammographies, par de petits points clairs.

### **I.5.2 Les masses :**

Les masses sont de grosses opacités opaques qui occupent de l'espace dans le sein et apparaissent sous forme de taches blanches sur les images de mammographie. Il peut s'agir d'un kyste (une collection de liquide non cancéreuse) ou d'une lésion solide, qui peut correspondre à un cancer du sein. Différentes caractéristiques permettent aux médecins de déterminer leur forme et densité par exemple.

### a- La forme :

Selon la description du BIRADS et de l'ACR [17], les masses mammaires peuvent avoir les formes suivantes (Figure I.5):

1. Ronde : masse circulaire ou sphérique.
2. Ovale : a une forme elliptique (ou ovoïde).
3. Lobulée : la forme de la masse présente une légère ondulation.
4. Irrégulière : cette appellation est réservée aux masses dont la forme est aléatoire.

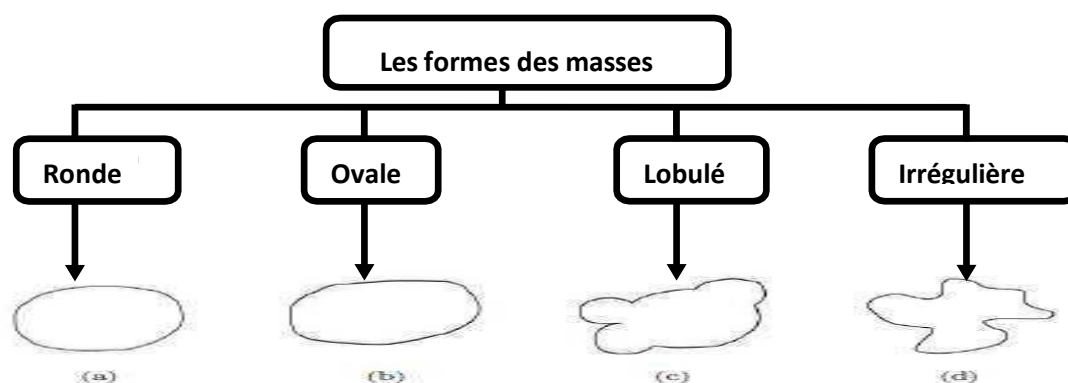


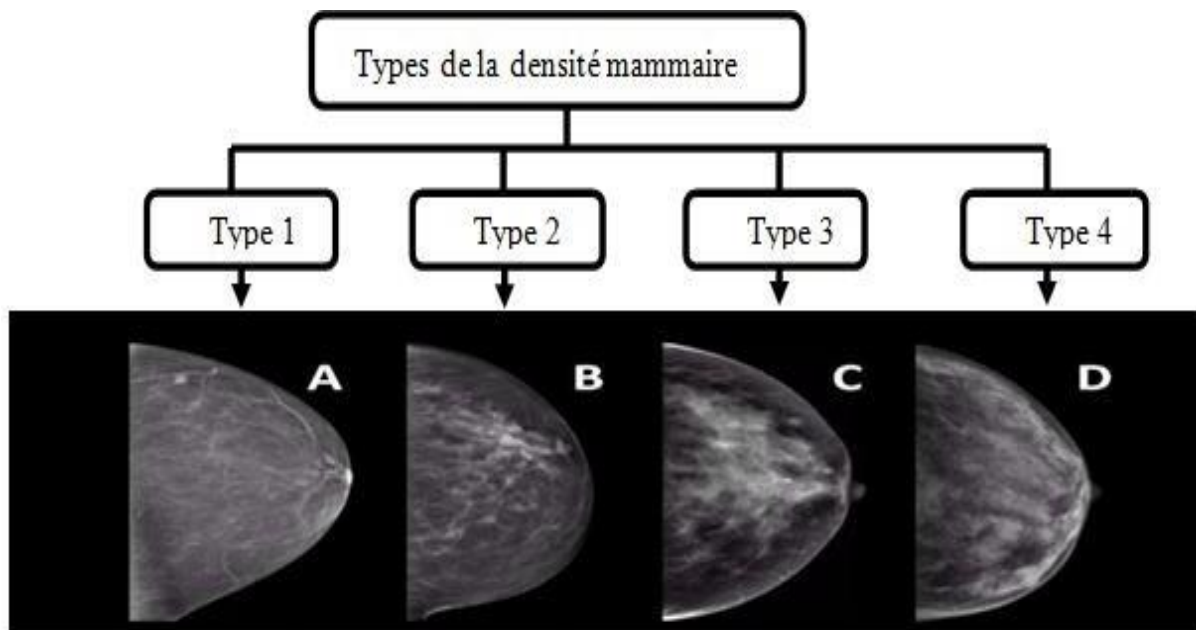
Figure I.5: Les différentes formes possibles d'une masse [18].

### b-La densité :

L'aspect du sein normal est très variable d'une femme à l'autre (voir figure 1.6). 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 [19], La densité du tissu mammaire est proportionnelle au risque de cancer. Afin de standardiser les comptes rendus des mammographies, la classification BIRADS (Breast Imaging Reporting And Data System) de l'ACR a défini quatre classes de la composition du sein [20] qui sont :

- Type 1 : sein presque entièrement gras : sein net avec densité < 25 %.
- Type 2 : il y a opacité fibreuse de densité entre 25 et 50 %.

- Type 3 : le tissu mammaire est dense et de densité hétérogène comprise entre 50 et 75 %.
- Type 4 : tissu mammaire extrêmement dense > 75 %.



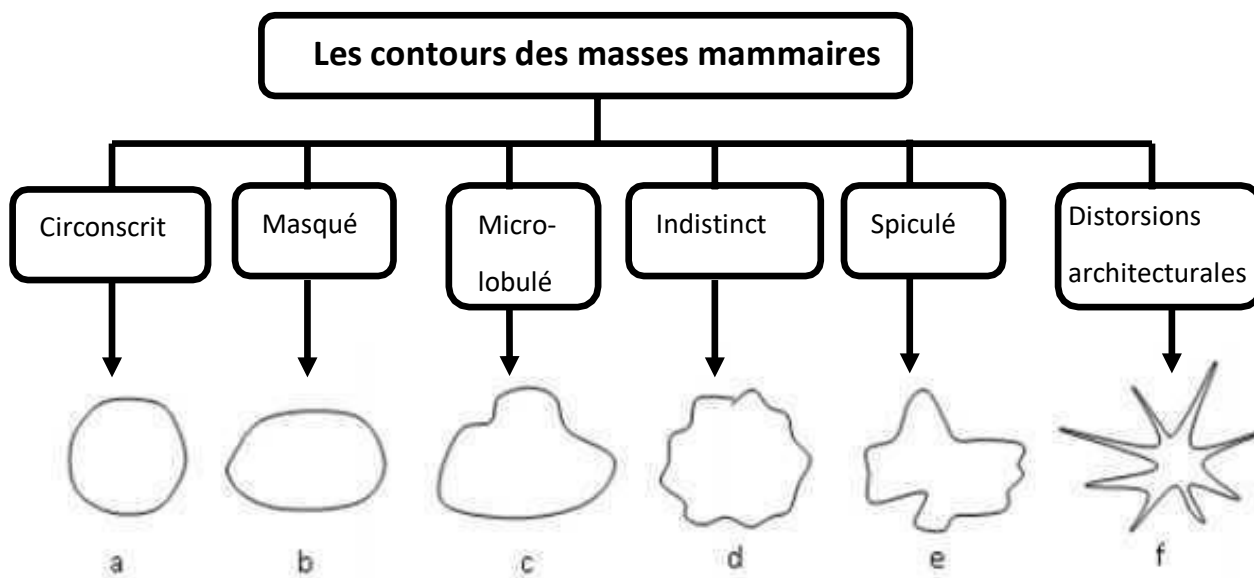
**Figure I.6:** Les différents types de densité mammaire [21].

#### **d-Le contour :**

Le BIRADS dénombre cinq classes de contour des masses mammaires (figure 1.7) :

1. Circonscrit : il s'agit d'une transition brusque entre la lésion et le tissu environnant.
2. Indistinct : dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre à une infiltration.
3. 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.
4. Spéculé : la masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.
5. Micro-lobulé : dans ce cas, de courtes dentelures du contour créent de petites ondulations.

6. Distorsions architecturales : ce sont des signes hautement suspects qui se traduisent par des structures linéaires ou des spécules qui convergent vers une même zone focale.



**Figure I.7:** Les différents contours d'une masse mammaire.[22]

## I.6 Diagnostic du cancer du sein assisté par ordinateur (DAOx)

Le diagnostic des lésions mammaires par les radiologues est une tâche subjective et des erreurs peuvent donc être commises, dont les conséquences peuvent être fatales pour la patiente. La combinaison de la fatigue, du manque d'expérience en imagerie mammaire et du nombre d'images à analyser chaque jour rendait cette tâche fastidieuse et justifiait les erreurs de diagnostic. En revanche, les systèmes automatisés de diagnostic assisté par ordinateur (CADx) conçus sur des modèles prenant en compte les descripteurs liés aux lésions mammaires démontreront systématiquement leur robustesse, ce qui contribuera à réduire le nombre de biopsies inutiles. En effet, il a été démontré que moins d'un tiers des mammographies suspectes après biopsie sont en réalité des mammographies de lésions cancéreuses [1]. De plus, les systèmes CADx sont conçus pour aider les radiologues à réduire le nombre de biopsies et à éviter le stress inutile du patient. [23] Ainsi, un système automatisé CADx utilisé dans un contexte de dépistage ou de diagnostic agirait comme un deuxième avis pour les radiologues et n'a absolument pas

Vocation à remplacer ces derniers. Typiquement, le fonctionnement d'un système CADx est une série d'étapes d'égale importance. Lorsque le système reçoit une image en entrée, il la prétraite d'abord, c'est-à-dire qu'il supprime le bruit d'acquisition ou le chatoiement et améliore le contraste de l'image pour obtenir la meilleure qualité d'image possible.

Cette étape est suivie de la détection d'anomalies (telles que des micro calcifications, des opacités ou des distorsions structurelles). Plusieurs attributs mathématiques sont ensuite extraits des anomalies détectées pour mieux décrire la nature de la lésion. Enfin, les descripteurs extraits sont classés selon un algorithme adapté à la tâche pour déterminer la nature de la qualité.

### **I.6.1 L'importance des systèmes de diagnostic par ordinateur :**

Même si une lésion est très suspecte à la mammographie ou à l'échographie, le diagnostic préopératoire est désormais indispensable pour deux raisons :

De nombreuses chirurgies inutiles pourraient être évitées s'il n'y avait pas le cancer.

En cas de cancer du sein, la chirurgie peut être optimisée d'emblée : si le cancer est de petite taille, par exemple, intervention au niveau des ganglions lymphatiques (par ex. ganglion sentinelle) peuvent être programmés et permettent au patient d'éviter le curage ganglionnaire Axillaire. Par exemple, dans d'autres cas, le traitement peut commencer par des médicaments (thérapie néo adjuvante) pour permettre une chirurgie moins sévère plus tard.

Dans de rares cas, ce diagnostic préopératoire n'est pas possible car la localisation radiographique, comme la biopsie, n'est pas disponible. La chirurgie permettra d'établir le diagnostic [24].

### **I.7 Les bases d'images de mammographie :**

Il existe différentes bases d'images utilisées par les chercheurs en imagerie du sein pour concevoir des systèmes automatiques de diagnostic assisté par ordinateur (DAOx). Parmi ces bases, on cite :

#### **I.7.1 Mammographie Image Analysis Society (MIAS)**

Mammographie Image Analysis Society (MIAS) [Suckling et al. 1994] est la plus ancienne base d'images publiques conçue au Royaume-Uni en 1994 et continue d'être largement utilisée dans l'état de l'art. La base MIAS contient 161 cas pour un total de 322 images numériques obtenues en incidence MLO, sur lesquelles sont présentes toutes les pathologies à savoir les

lésions bénignes et malignes, mais aussi les images normales. Cette base contient un grand nombre de masses spéculées, ainsi que des informations sur la densité mammaire ; Cependant, la classification des masses effectuée par les radiologues qui ont évalué cette base ne répond pas aux normes ACR [25]. Cependant, avec l'utilisation accrue de la classification ACR BI-RADS, certains auteurs ont tenté de classer les masses MIAS afin qu'ils puissent répondre aux normes actuelles d'imagerie mammaire. Un autre défaut de la base MIAS est l'explication de ces images, à savoir la référence au centre et au diamètre de la zone d'intérêt, i.e. la zone où la pathologie est présente.

### **I.7.2 Digital Database For Screening Mammography (DDSM)**

La base de données de mammographie numérique (DDSM), créée en 1999 aux États-Unis d'Amérique, est la plus grande base de données publique de mammographie. Il contient 2620 Cas dont deux images (MLO et CC) par sein, pour un total de 10480 images où toutes les maladies ont été trouvées, des images ordinaires à celles contenant des grosseurs bénignes et malignes. Pour chacune de ces images, l'âge de la patiente, la densité mammaire et la classification BI-RADS sont fournis.

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 [26].

### **I.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 [27].

## **I.8 Conclusion**

Dans ce chapitre, nous avons introduit quelques notions concernant l'anatomie du sein, les pathologies qui le touchent, les différentes modalités d'imagerie mammaire et les systèmes d'aide au diagnostic. Dans le prochain chapitre, nous allons voir les différentes étapes pour le traitement des images mammographies, qui permettent d'améliorer les performances du classifieur.

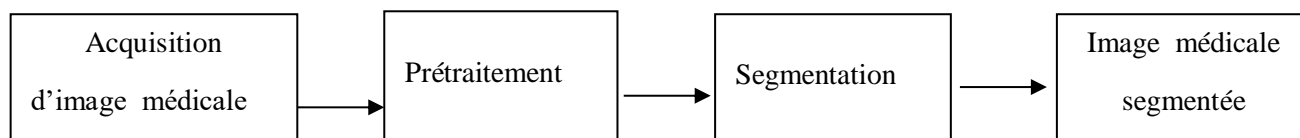
# **CHAPITRE II : Traitement des Images de Mammographies**



## II.1 Introduction

Les mammographies sont des images qui n'offrent pas un bon contraste entre les tissus mammaires sains et malins, car l'atténuation des rayons X varie peu entre ces tissus, en particulier chez les jeunes femmes présentant un parenchyme mammaire très dense. Le traitement de l'image mammographique fait en deux phases (Figure II.1) :

- Une phase de prétraitement comprenant 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.
- Une phase de segmentation qui consiste à diviser une image en sous-ensembles appelés régions. La segmentation est alors obtenue en extrayant des composantes connexes de pixels appartenant à une même classe. Par conséquent, une région sera constituée de pixels de la même classe.



**Figure II.1**–Traitement d'image mammographique.

## II.2 Traitement d'image mammographique

### II.2.1 Prétraitement des images mammographies

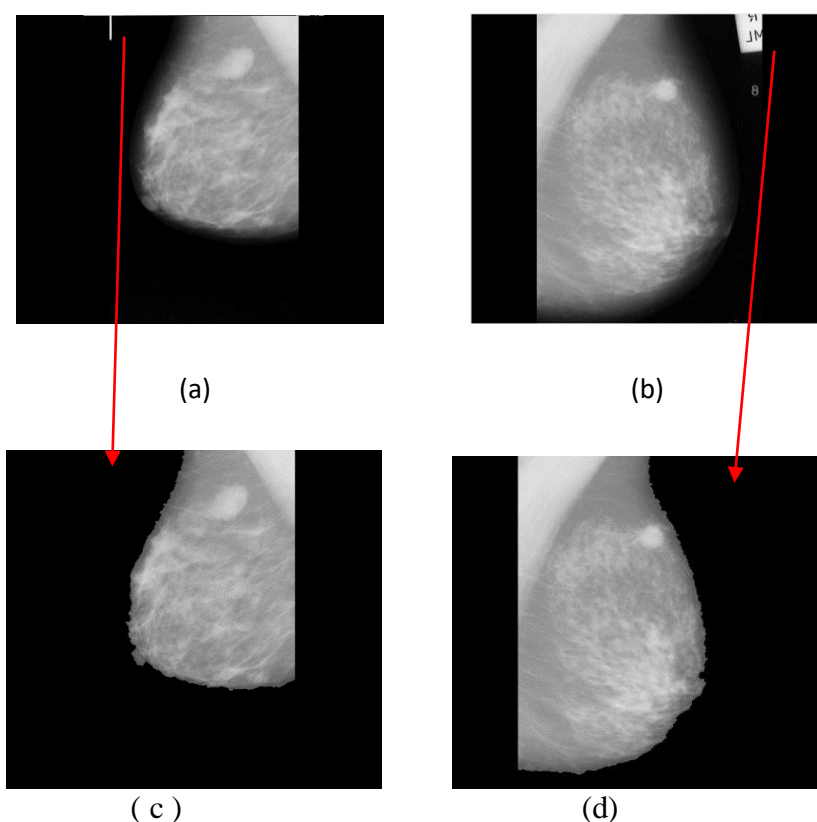
Les techniques standard de réduction du bruit et d'amélioration du contraste ont d'abord été appliquées aux mammographies avant de se rendre compte qu'en plus d'améliorer l'image, ces méthodes contribuaient également à amplifier de manière significative le bruit présent dans l'image. Il suggère ensuite des méthodes plus spécifiques qui aident souvent à améliorer la qualité des mammographies, par exemple la note indique également que l'utilisation de ces techniques dans un cadre clinique pourrait réduire de moitié la dose de rayonnement lorsque les mammographies sont enregistrées. Cependant, le vrai problème avec les algorithmes de prétraitement est que certaines régions peuvent être bien traitées, tandis que

d'autres peuvent être traitées plus ou moins bien. 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 [29]. Les techniques d'amélioration d'image offrent de nombreuses possibilités d'améliorer la qualité visuelle.

Ces techniques sont largement limitées par les détails de l'image et les types d'informations qui doivent être améliorées et/ou éliminées. Cette étape vise à améliorer la qualité de l'image traitée en appliquant d'abord un filtre médian pour réduire le bruit tout en préservant les bords de l'image, suivi d'un filtre morphologique pour supprimer les intensités dans l'image. Par conséquent, un bon résultat est que le choix correct de ces techniques est fortement influencé par l'importance du détail de l'image et le type d'information que l'on souhaite améliorer et/ou éliminer. [30]

### **II.2.1.1 Suppression des artefacts**

Les images mammographies numériques contiennent généralement deux zones distinctes : la zone exposée du sein et une zone non exposée, le fond. Ce fond peut contenir des artefacts qui sont facilement ignorés par le système visuel humain lors de l'interprétation. En revanche, un système automatisé doit, d'abord, identifier et classer ces artefacts qui engendrent des erreurs d'interprétation en analysant les images utilisées (Figure II.2). Les mammographies sont généralement marquées d'une sorte d'étiquette d'identification permanente qui contient des informations sur le test effectué. Ces étiquettes sont des indicateurs radio-opaques qui montrent la latéralité mammographie (R/L, droite/gauche) et des indicateurs de projection MLO/CC pour l'axe X de la mammographie [ 31].



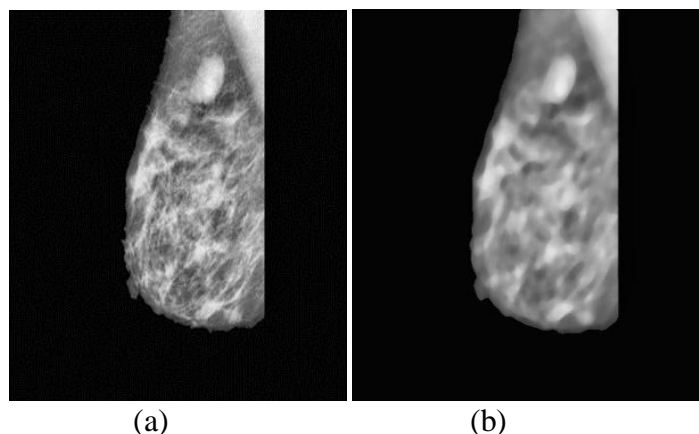
**Figure II.2** Exemple Suppression des artefacts : (a) et (b) images originelles, (c) et (d) après élimination des étiquettes.

### II.2.1.2 Réduction du bruit

La réduction du bruit est obtenue en appliquant un filtre médian pour éliminer le bruit causé par le dispositif d'acquisition.

#### II.2.1.2. 1 Filtre médian

Le filtre médian est un filtre numérique non linéaire couramment utilisé pour la réduction du bruit en préservant les contours des images [32]. L'idée principale du filtre médian est de remplacer chaque entrée par la valeur médiane de son voisinage. (Figure II.3)

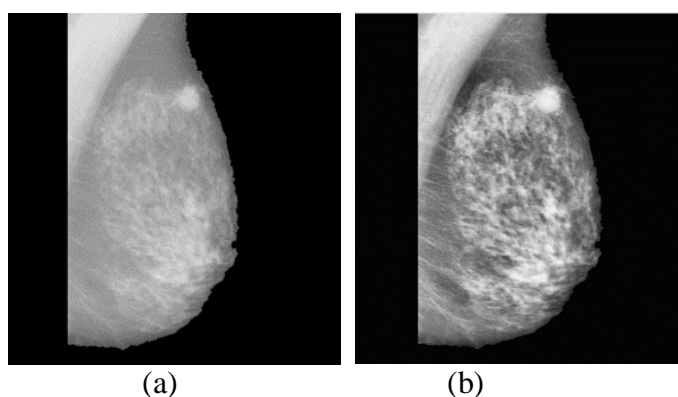


**Figure II.3** : Suppression du bruit : (a) d'image obtenue précédemment, (b) d'image filtrée.

### II.2.1.2. 2 Égalisation adaptative d'histogramme

Pour améliorer les caractéristiques d'une image (en général, pour accentuer le contraste), il est courant d'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 est d'appliquer une égalisation d'histogramme à chaque pixel et à sa zone de contexte. Cette région représente en effet les pixels voisins autour du pixel traité [33]. (Figure II.4)



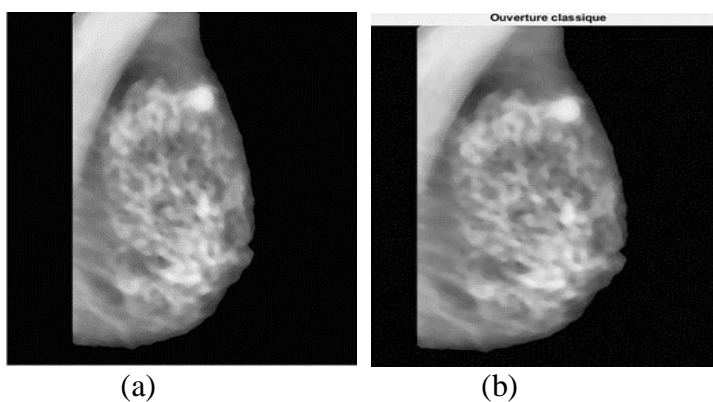
**Figure II.4-** Exemple d'image améliorée du contraste. (a) d'image après la suppression des raies et des étiquettes, (b) amélioration de la qualité d'image (contraste) par égalisation adaptative d'histogramme.

### II.2.1.2. 3. Amélioration de contraste

Le contraste dans une image peut être défini comme étant l'écart type de la variation du niveau de gris dans l'image. Dans le cadre des images médicales, ce traitement vise à mettre en évidence des régions d'intérêt tout en en atténuant d'autres pour faciliter le diagnostic. En fait dans cette partie nous présentons amélioration du contraste par :

#### a) Ouverture morphologique classique

L'ouverture est un type de filtre morphologique couramment utilisé pour restaurer ou restaurer l'image d'origine autant que possible. L'ouverture morphologique est égale à la combinaison d'une première érosion puis d'une expansion. C'est un opérateur incrémental et idempotent, c'est-à-dire que le résultat ne change pas si l'opérateur est appliqué plusieurs fois, il suffit de l'appliquer une seule fois. Sa fonction est de faire disparaître les petites particules et de séparer les grosses particules là où elles sont plus fines. (Figure II.5)



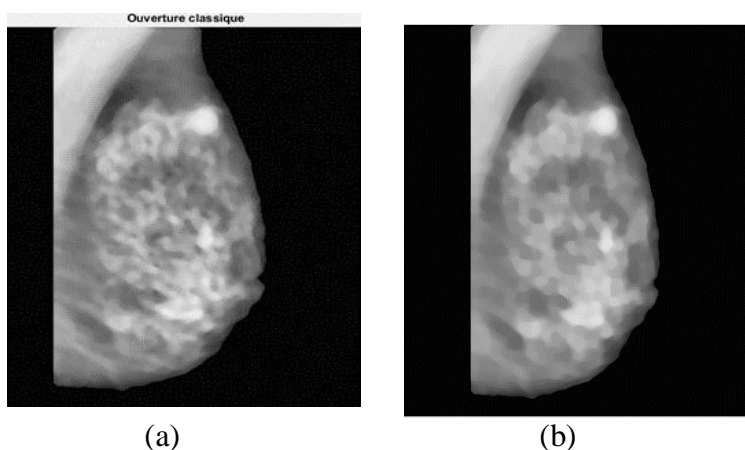
**Figure II.5-** Amélioration du contraste. (a) d'image filtrée, (b) rehaussement de contraste par une ouverture.

#### b) Filtres Alternés Séquentiels (FAS)

À partir des filtres morphologiques primaires, tels que les ouvertures et les fermetures, des filtres morphologiques plus avancés sont identifiés. Parmi ceux-ci, les filtres séquencés en

alternance (FAS) sont probablement les plus courants. Il réduit la complexité de la magie sans insulter les contradictions les plus importantes.

Aperture supprime les détails blancs sans altérer les parties sombres de l'image, tandis que Close fait le contraire. Les filtres alternatifs séquentiels (FAS) adoptent un comportement plus symétrique en alternant des combinaisons d'activation et de désactivation. [34] (Figure II.6)



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

**Figure II.6-** Application du Filtre Alterné Séquentiel sur une image mammographie.

### II.3 Segmentation des images mammographie

La segmentation des images mammographies fait référence à la division d'une image mammographie en différentes régions ou structures d'intérêt. Cela implique d'identifier et de séparer les régions d'intérêt telles que les tumeurs, les masses, les micro calcifications ou les autres anomalies présentes dans l'image.

La segmentation joue un rôle crucial dans l'analyse et le diagnostic des images mammographies. En isolant et en mettant en évidence les structures d'intérêt, la segmentation facilite la détection des anomalies et la caractérisation des lésions, ce qui peut contribuer à la détection précoce du cancer du sein.

Différentes méthodes peuvent être utilisées pour segmenter les images mammographies, allant des techniques traditionnelles basées sur le seuillage et les contours aux méthodes plus

avancées basées sur l'apprentissage automatique, telles que les réseaux de neurones convolutés. L'objectif principal de la segmentation est de fournir des régions précises et fiables pour faciliter

l'analyse et l'interprétation des images mammographies par les radiologues ou les systèmes d'aide au diagnostic. [35]

### **II.3.1 Algorithme de k-Means**

La segmentation par k-Means est un algorithme de clustering largement utilisé dans le domaine de l'apprentissage automatique non supervisé. L'algorithme renvoie une partition d'une image dans laquelle les individus de chaque cluster sont plus proches les uns des autres et plus loin des individus des autres clusters, chaque cluster de la partition est défini par ses individus et son centre. Il minimise la somme des distances entre chaque individu et le centre de son cluster. [36] (Figure II.7)

Les étapes de l'algorithme K\_Means sont décrites comme suit :

- 1) choisit k points au hasard, ces points sont les centres.
- 2) calculer la matrice de distance entre les individus et les k centres, ont obtenu k groupes où chaque groupe a un centre et des individus les plus proches de ce centre.
- 3) calculer pour chaque groupe les nouveaux centres.
- 4) calculer la matrice de distance entre les individus et les k nouveaux centres.
- 5) répéter les étapes 3 et 4.

**Entrée**

Ensemble de N données, noté par x

Nombre de groupes souhaité, noté par k

**Sortie**

Une partition de k groupes  $\{C_1, C_2, \dots, C_k\}$

**Début**

- 1) Initialisation aléatoire des centres  $C_k$  ;

**Répéter**

- 2) Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche :

$$x_i \in C_k \text{ si } \forall j \quad |x_i - \mu_k| = \min_j |x_i - \mu_j| \quad \dots(\text{II.1})$$

Avec  $\mu_k$  le centre de la classe K ;

- 3) Représentation : Calculer les centres associée à la nouvelle partition ;

$$\mu_k = \frac{1}{N} \sum_{x \in C_k} x_i \quad \dots(\text{II.2})$$

**Fin**

**Figure II.7-** Organigramme de méthode de K\_Means.



## **II.4 Conclusion :**

Les étapes de traitement d'image dans le 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 ainsi que le principe de la segmentation par k-means.

**CHAPITRE III :**  
**Les machines à Vecteur Support**  
**( SVM)**

### III.1 Introduction

Support Vector Machine (SVM) également appelé « séparateurs à vaste marge ou les machines à vecteurs de support » est une nouvelle approche de la classification de modèles supervisés qui a été appliquée avec succès à un large éventail de problèmes de reconnaissance de modèles. Elle constitue également un algorithme pour l'apprentissage des règles de classification et de régression à partir de données. Le modèle SVM est le plus approprié pour travailler avec précision et efficacité avec des espaces de fonctions de haute dimensionnalité. Le SVM est basé sur des bases mathématiques solides et donne des résultats simples et des algorithmes très puissants.

L'algorithme SVM a été développé dans les années 90 par le russe Vladimir Vapnik [37]. Initialement, les SVM ont été développés comme un algorithme de classification binaire supervisée. Il s'avère particulièrement efficace de par le fait qu'il peut traiter des problèmes mettant en jeu de grands nombres de descripteurs, qu'il assure une solution unique (pas de problèmes de minimum local comme pour les réseaux de neurones) et il a fourni de bons résultats sur des problèmes réels.[38]

### III.2 Support Vector machine (SVM)

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

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

Le premier consiste à définir l'hyperplan comme solution d'un problème d'optimisation sous contraintes dont la fonction objective ne s'exprime qu'à l'aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes actives ou vecteurs supports contrôle la complexité du modèle.

Le passage à la recherche de surfaces séparatrices non linéaires est obtenu par l'introduction d'une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (feature space) de plus grande dimension. D'où l'appellation couramment rencontrée de machine à noyau ou kernel machine.[39]

### III.3 Principe de fonctionnement

Soit un ensemble d'exemples d'apprentissage  $(X_i, Y_i)$  où les  $X_i$  sont les instances de données du mode réel et les  $Y_i$  les étiquettes indiquant la classe à laquelle appartient l'instance.

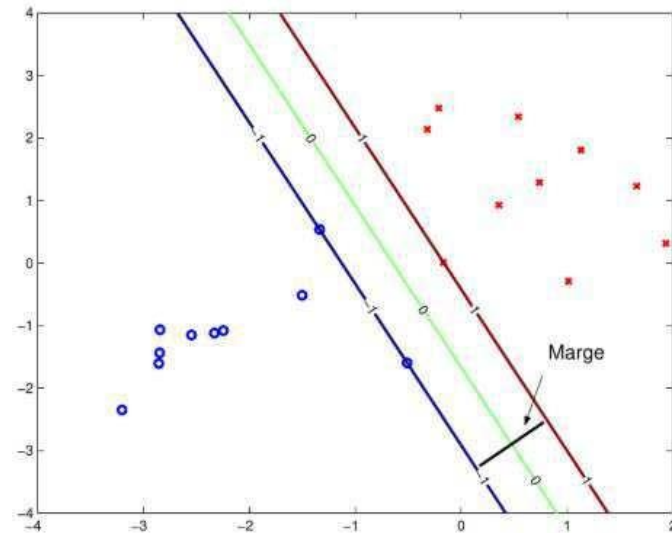
Pour le problème de reconnaissance de modèle à deux classes,  $Y_i = +1$  ou  $Y_i = -1$ . Un exemple d'apprentissage  $(X_i, Y_i)$  est appelé positif si  $Y_i = +1$  et négatif si  $Y_i = -1$ . On cherche alors une « surface de séparation »  $f : X \rightarrow \mathbb{R}$  tel que si  $f(x) > 0$  alors  $X$  est affecté à la classe  $+1$  et si  $f(x) < 0$  alors  $X$  est affecté à la classe  $-1$ .

Pour un problème de classification linéaire on suppose que les deux classes ( $-1$  et  $+1$ ) sont séparable par un hyperplan, la fonction  $f$  a donc la forme :

$$F(x) = \sum_{i=1}^n W_i X_i + b \quad (3.1)$$

Où  $w$  est le vecteur orthogonal à l'hyperplan et  $b$  est le déplacement par rapport à l'origine. Pour juger la qualité d'un hyperplan en tant que séparateur on utilise la distance entre les exemples d'apprentissage et ce séparateur. Plus précisément, la « marge » d'un problème d'apprentissage est définie comme la distance entre le plus proche exemple d'apprentissage et l'hyperplan de séparation. Pour un hyperplan  $H$  on a [40] (Figure 3.1) :

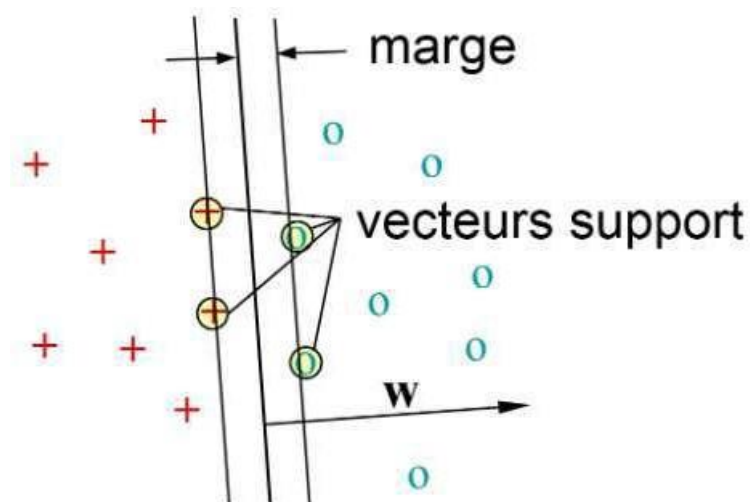
$$Marge(H) = \min d(x_i, H) \quad (3.2)$$



**Figure III.1** – Le séparateur idéal correspond intuitivement à l’hyperplan qui passe au milieu, c’est le séparateur de marge maximale [41].

Les SVM linéaires cherchent le séparateur (l’hyperplan de séparation) qui maximise la marge. On appelle cela séparateur à vaste marge.

Les éléments de la classe 1 les plus proches de ce séparateur se trouvent à la même distance du séparateur que les éléments les plus proches de la classe 2 (cette distance est égale à la marge). Ces éléments, soit d’une classe soit de l’autre, s’appellent vecteurs de support.



**Figure III.2** – Les vecteurs de support se trouvent à une distance égale à la marge d’un côté ou de l’autre de l’hyperplan de séparation [42].

Intuitivement, ce sont les vecteurs de support qui déterminent le séparateur (par l'intermédiaire de la fonction distance et de leur configuration géométrique). Une fois le séparateur  $f(x)$  trouvé, la classification d'un nouvel exemple se fait par une simple décision à seuil zéro :

$f(x) = 0$  : l'élément se trouve sur la frontière de séparation, pas de décision.

$f(x) > 0$  : classe 1.

$f(x) < 0$  : classe 0.

### III.4 Formalisme des SVM

Fondamentalement, un SVM cherchera simplement à trouver un hyperplan qui sépare les deux catégories d'un problème. Soit l'espace vectoriel  $E$  de dimension  $n$ , doté de la base  $(e_1, \dots, e_n)$ . L'équation caractéristique d'un hyperplan est de la forme :

$$w_1.e_1 + w_2.e_2 + \dots + w_n.e_n = 0 \quad (3.3)$$

Ainsi, si l'on se place dans  $\mathcal{R}^n$  muni de la base  $(e_1, \dots, e_n)$ , pendant son entraînement le SVM calculera un hyperplan vectoriel d'équation  $w_1.e_1 + w_2.e_2 + \dots + w_n.e_n = 0$ , ainsi qu'un scalaire (un nombre réel)  $b$ . c'est ce scalaire  $b$  qui va nous permettre de travailler

Avec un hyperplan affine. Le vecteur  $\begin{pmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_n \end{pmatrix}$  est appelé vecteur de poids, le scalaire  $b$  est appelé

biais.

Une fois l'entraînement terminé, pour classe une nouvelle entrée  $x = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$  le SVM regardera

le signe de :

$$F(x) = w_1 a_1 + w_2 a_2 + \dots + w_n a_n + b = \sum_{i=1}^n w_i a_i + b = w^t \cdot a + b$$

Si  $f(x)$  est positif ou nul, alors  $X$  est d'un coté de l'hyperplan affine et appartient à la première catégorie, sinon  $X$  appartient à la seconde catégorie. [43]

### III.5 Entraînement des SVM

En conséquence, étant donné un hyperplan de vecteur de poids  $w$ , et de biais  $b$ , nous pouvons calculer si un point  $x_k$  appartient à telle ou telle catégorie, grâce au signe de  $f(x_k)$ . Supposons que l'on assigne à tout point  $x_k$  un label  $l_k$  qui vaut 1 si  $x_k$  appartient à la première catégorie, et -1 si  $x_k$  appartient à la seconde catégorie. Alors, si le SVM est correctement entraîné, on a toujours  $l_k f(x_k) \geq 0$ , c'est-à-dire  $l_k (w^t \cdot x_k + b) \geq 0$ . Le but d'un SVM, lors de l'entraînement, est donc de trouver un vecteur de poids  $w$  et un biais tels que, pour tout point  $x_k$  de label  $l_k$  appartenant aux données d'entraînement  $l_k (w^t \cdot x_k + b) \geq 0$

Autrement dit, de trouver un hyperplan séparateur entre les deux catégories. On Choisira l'hyperplan qui maximise la marge. [44]

#### III.5.1 Calcul de la marge

La distance d'un point  $x_k$  à l'hyperplan de vecteur poids  $w$  et de biais  $b$  est donnée par :

$$\frac{l_k (w^t \cdot x_k + b)}{\|w\|} \quad (3.4)$$

Où  $\|w\|$  désigne la norme euclidienne de  $w$ . la marge (distance minimale de l'hyperplan à un des points d'entraînement) d'un hyperplan de paramétré  $(w,b)$  par rapport à un ensemble de points  $x_k$  est donc : [45]

$$\min_k \frac{l_k (w^t \cdot x_k + b)}{\|w\|} \quad (3.5)$$

### III.5.2 Maximisation de la marge

L'objectif est de trouver l'hyperplan qui permet de maximiser la marge. Autrement dit, chercher l'unique hyperplan dont les paramètres  $(w, b)$  sont donnés par la formule :

$$\text{Argmax } (w,b) \frac{\min_k (l_k w^t \cdot x_k + b)}{\|w\|} \quad (3.6)$$

Il existe plusieurs couples  $(w, b)$  qui décrivent le même hyperplan. On décide de ne considérer que l'unique paramétrage  $(w, b)$  tel que les vecteurs support  $x_s$  vérifient  $l_s (w^t \cdot x_s + b) = 1$  par conséquent  $l_k (w^t \cdot x_k + b) \geq 1$ , et l'égalité est atteinte si  $x_k$  est un vecteur support. Dit autrement, cette normalisation sur  $w$  et  $b$  permet de garantir que la marge citée dans l'équation 3.6 est alors de :

$$\text{mink} \frac{l_k (w^t \cdot x_k + b)}{\|w\|} \quad (3.7)$$

Le problème d'optimisation de la marge 3.7 se simplifie ainsi en :

$$\text{argmax}(w,b) \frac{\text{mink}(l_k w^t \cdot x_k + b)}{\|w\|} \quad (3.8)$$

Tout en gardant d'hypothèses de normalisation, à savoir :

$$\forall_k l_k (w^t \cdot x_k + b) \geq 1 \quad (3.9)$$

Le problème revient à :

$$\begin{cases} \text{maximiser} & \frac{1}{\|w\|} \\ \text{sous la contrainte} & \forall_k l_k (w^t \cdot x_k + b) \geq 1 \end{cases} \quad (3.10)$$



Que l'on peut reformuler pour des raisons pratiques de la façon suivante :

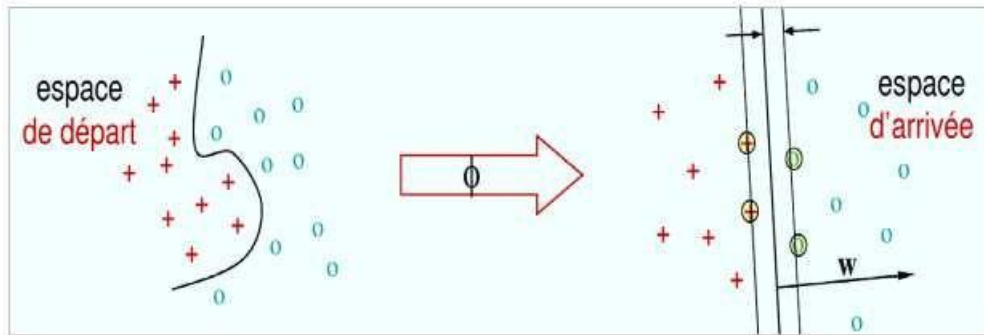
$$\begin{cases} \text{minimiser} & \|w\|^2 \\ \text{sous la contrainte} & \forall_k l_k(w^t \cdot x_k + b) \geq 1 \end{cases} \quad (3.11)$$

Ce genre de problème est appelé problème d'optimisation quadratique, et il existe de nombreuses méthodes pour le résoudre. Dans le cas présent, on utilise la méthode des multiplicateurs de Lagrange. Cette résolution donnera une valeur optimale pour  $w$ . Pour retrouver  $b$ , il faut résoudre l'équation suivante : [46]

$$\text{Min}_k l_k(w^t \cdot x_k + b) = 1 \quad (3.12)$$

### III.6 SVM non linéaire et la fonction noyau

Dans ce paragraphe, le problème des données non linéairement séparables est présente. Pour rappel, des données sont non linéairement séparables quand il n'existe pas d'hyperplan capable de séparer correctement les deux catégories. Ce qui, d'ailleurs, arrive quasiment tout le temps en pratique. Afin de remédier à ce problème, une solution consiste à transposer les données dans un autre espace (en général de plus grande dimension) dans lequel elles sont linéairement séparables (ou presque) et ensuite appliquer l'algorithme SVM sur les données transposées. 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 nouvel espace de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau [47].

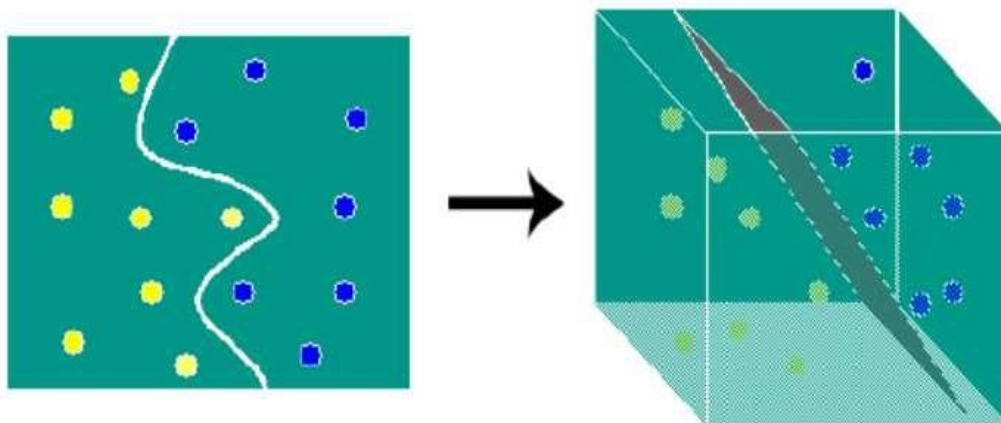


**Figure III.3** – Astuce à noyaux : projeter les données dans un espace de dimension beaucoup plus grande, où elles deviennent séparables linéairement [48].

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et Laplace.

### III.6.1 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 représentation de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau.



**Figure III.4** – Exemple de changement de l'espace de données [49].

## **I.7 Conclusion**

Dans ce chapitre, nous avons donné une vision générale des SVM, nous avons essayé de présenter de manière simple et complète le concept de systèmes d'apprentissage introduit par Vladimir Vapnik. Cette méthode de classification est basée sur la recherche d'un hyperplan, ce qui permet de séparer au mieux les jeux de données.

Les machines à vecteurs de support sont un outil qui s'est récemment étendu à tous les niveaux. Leur principale force réside dans leurs capacités de généralisation et d'optimisation, qui se font sans minima locaux et sans phénomènes de sur-apprentissage, contrairement à d'autres outils plus classiquement utilisés (réseaux de neurones, ...) autrement, ce qui en fait un outil de développement intéressant.

**CHAPITRE IV :**  
**Conception et Implémentation**  
**d'un Système d'aide au Diagnostic**  
**assisté par Ordinateur (CAOx)**

### **IV.1. Problématique**

Comme le nombre de mammographies a augmenté ces dernières années, il demeure difficile pour les radiologues de fournir une analyse précise et cohérente, d'où l'importance de l'aide au diagnostic développée au cours des deux dernières décennies. La Résolution Support System (CAOx) pour l'imagerie médicale est un programme numérique qui aide les experts à interpréter et à classer les images médicales. Son rôle est de détecter une ou plusieurs tumeurs sur l'image et de les classer comme malignes ou bénignes.

L'objectif de ce projet est de mettre au point un système d'aide semi-automatique à la détection et la classification des différentes pathologies dans des images mammographiques. Le système qu'on propose se voit divisé en quatre étapes que sont le prétraitement, la segmentation, l'extraction des descripteurs et enfin la classification qui fera l'aide des SVM.

### **IV.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 MIAS. Les images sont MLO (lésion latérale inclinée) et chaque image à une résolution spatiale de 1024 1024 pixels. Les différentes maladies des mammifères dans la base de données MIAS, représentent principalement des blocs ronds, des spéculateurs et des distorsions architecturales bénignes et malveillantes. Notre base de données continent 41 images distribuées comme suit:

- (19) images de masses malignes.
- (22) images de masses bénignes.

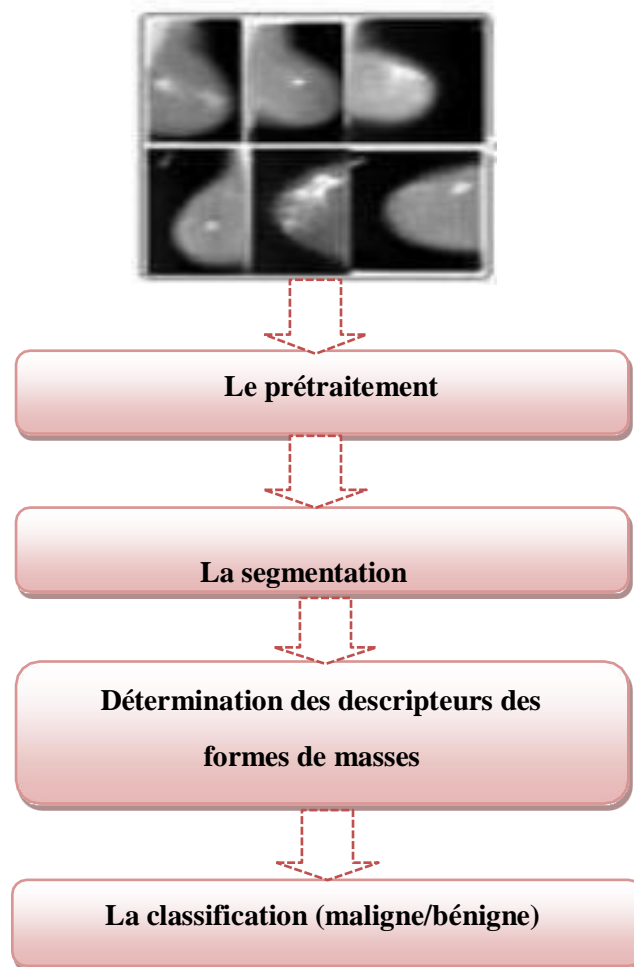
### **IV.3 Conception du système d'aide au diagnostic assisté par ordinateur**

La conception du système de soutien diagnostique 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.

## Chapitre04 : Conception du système d'aide au diagnostic assisté par ordinateur

Pour ce faire, nous suivons les étapes décrites à la figure IV.1.



**Figure IV.1-** les étapes proposées de la conception du système d'aide au diagnostic.

Dans ce qui suit, nous allons détailler ses différentes étapes comme indiqué dans la figure IV.2:

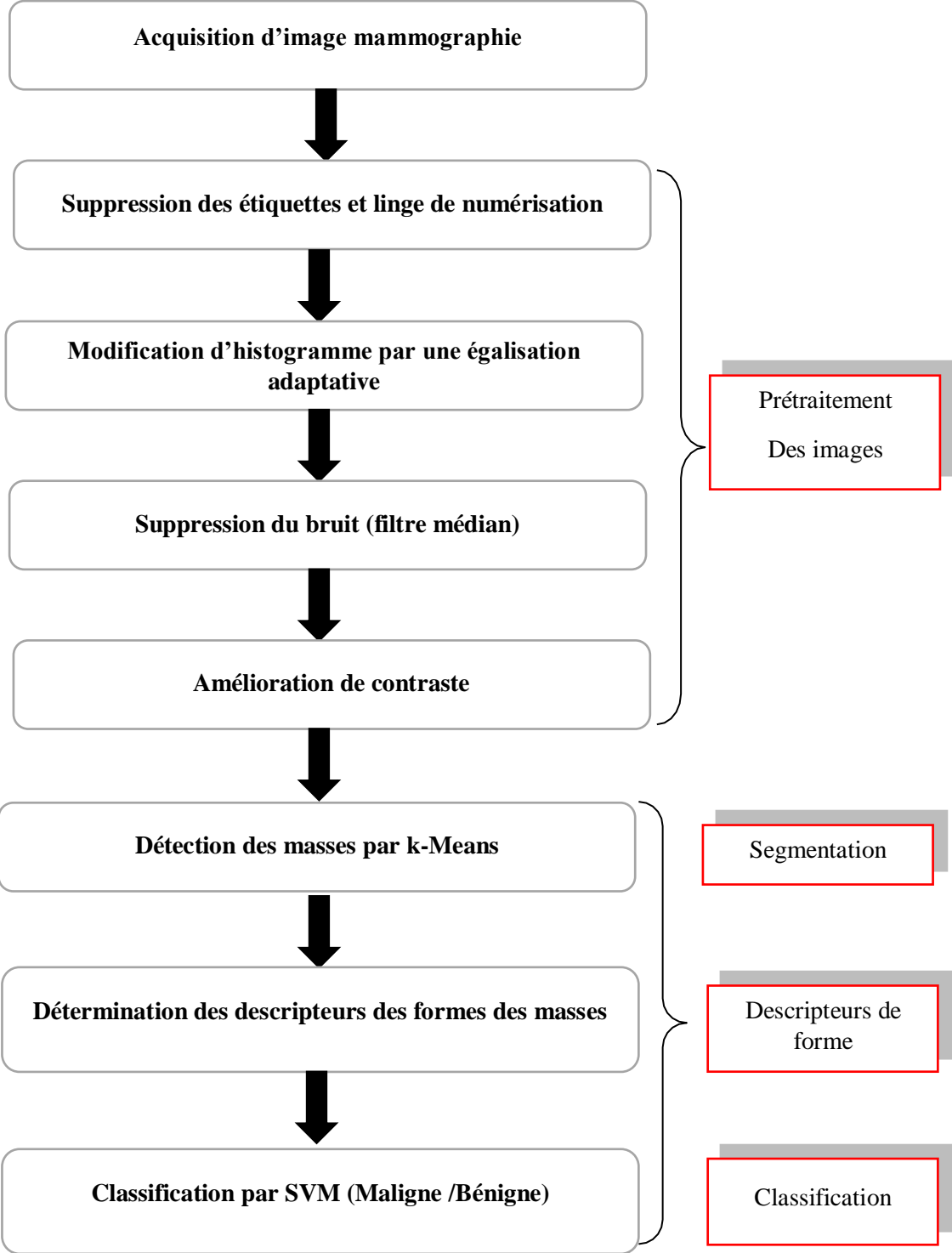


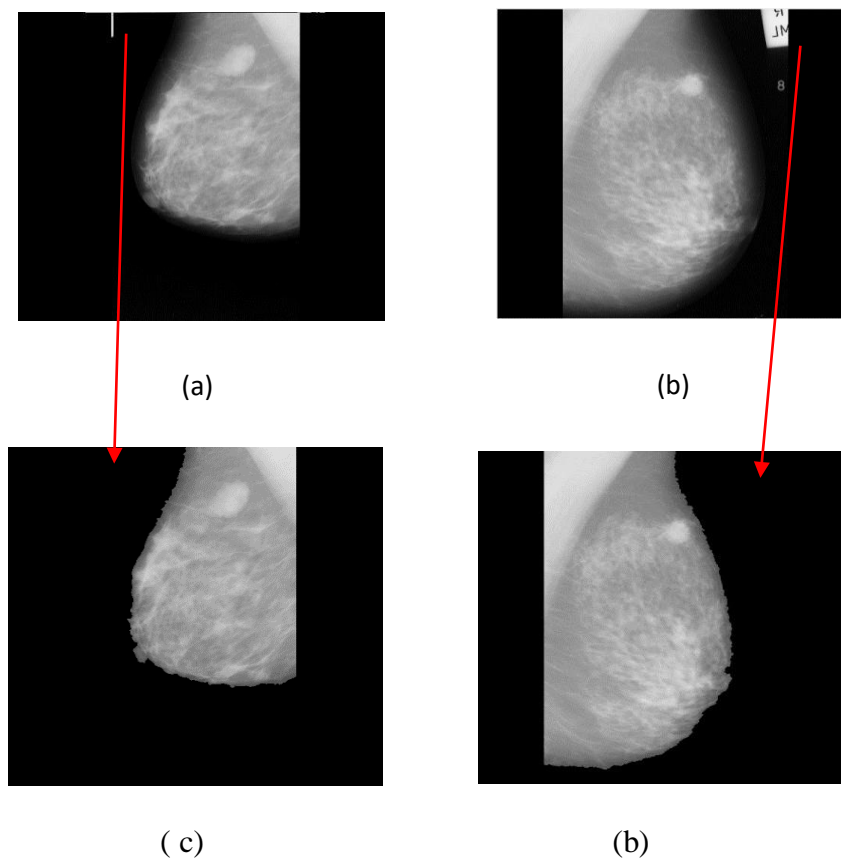
Figure IV.2 : Les différentes étapes du traitement des mammographies.

### IV.3. 1.Prétraitement des données

Les bruits de capteurs et de lumière contaminent fréquemment les images acquises, qu'elles soient médicales ou non. Pour cela, une étape de prétraitement est nécessaire pour chaque analyse d'image.

#### A.Suppression des lignes et des étiquettes de numérisation

Nous avons utilisé deux fonctions pour détecter et supprimer les lignes de balayage : seuillage et suppression des petites régions. Nous avons souillé tout ce qui ne fait pas partie du sein sur la mammographie d'origine est automatiquement supprimé. Un exemple de la base de données mini-MIAS est présenté à la figure IV.3.

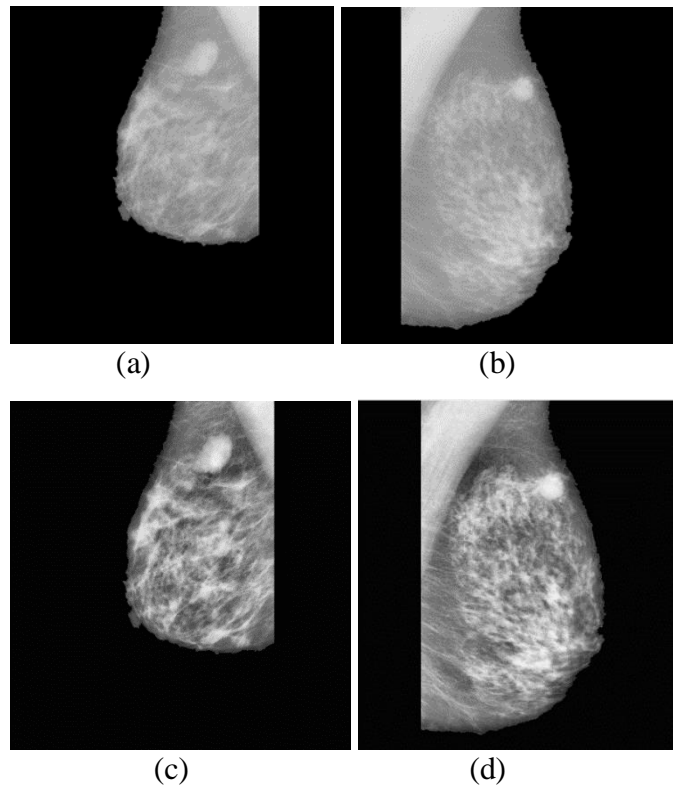


**Figure IV.3** : Suppression des artefacts : (a) et (b) images originelles, (c) et (d) après élimination des étiquettes.



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

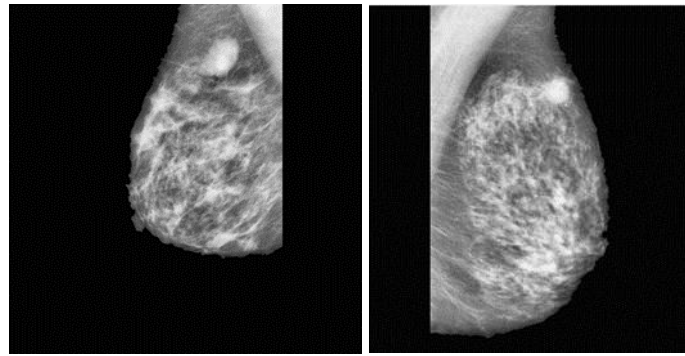
Nous avons amélioré le contraste en utilisant l'égalisation adaptative d'histogramme avec une implémentation sous Matlab. (Figure IV.4):



**Figure IV.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.

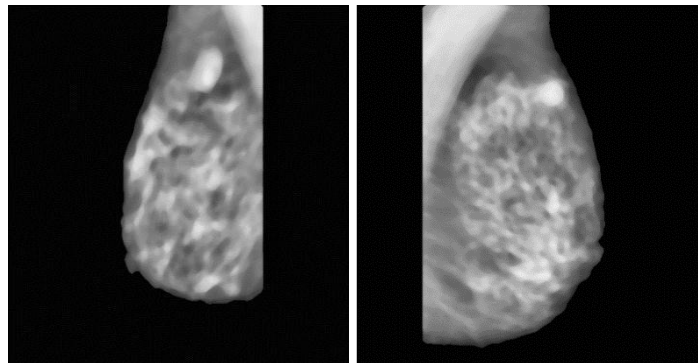
### B. Filtrage

Nous avons utilisé un filtre médian pour éliminer le bruit. Les résultats obtenus sont présentés dans la figure IV.5



(a)

(b)



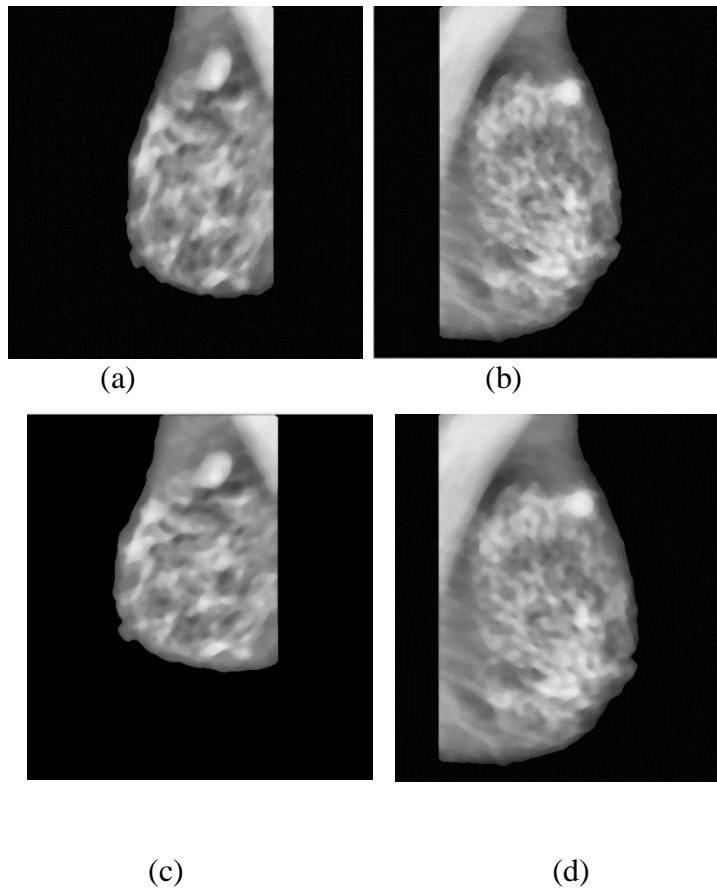
(c)

(d)

**Figure IV.5** -Suppression du bruit : (a) et (b) images obtenues précédemment, (c) et (d) images filtrées.

### C. Rehaussement de contraste

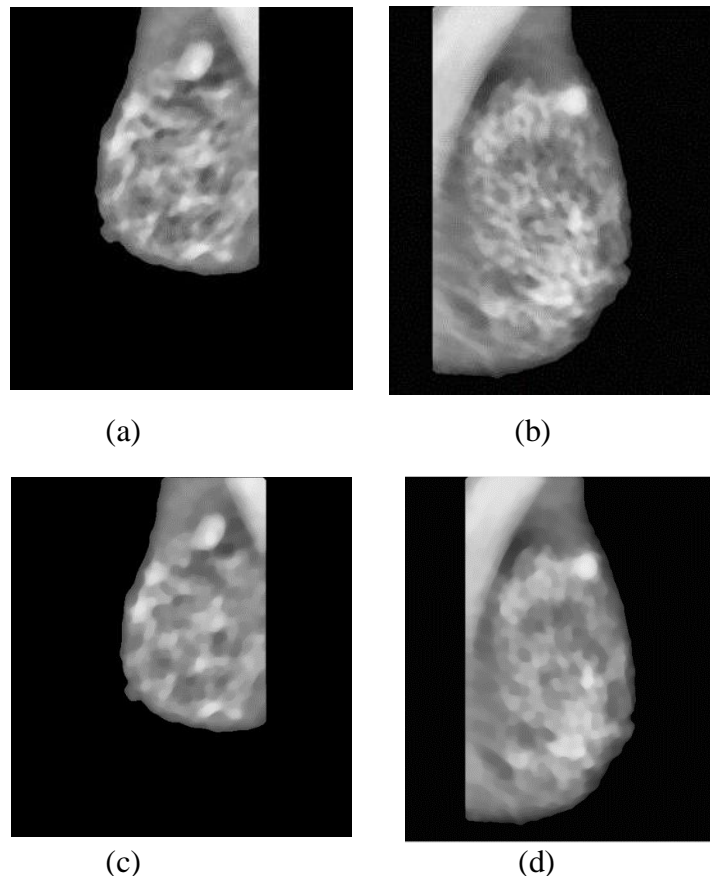
Après le filtrage de l'image, un rehaussement de contraste est effectué pour mettre en évidence toutes les taches de haute intensité, autrement toutes les régions susceptibles d'être des masses. Dans un premier temps, nous avons amélioré le contraste en utilisant un filtre morphologique (voir chapitre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



**Figure IV.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 (Filtres alternés séquentiels) (voir chapitre II). 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. (Figure IV.7):



**Figure IV.7 :** Amélioration du contraste : (a) et (b) images filtrées par ouverture, (c) et (d) images avec rehaussement de contraste par filtre FAS.

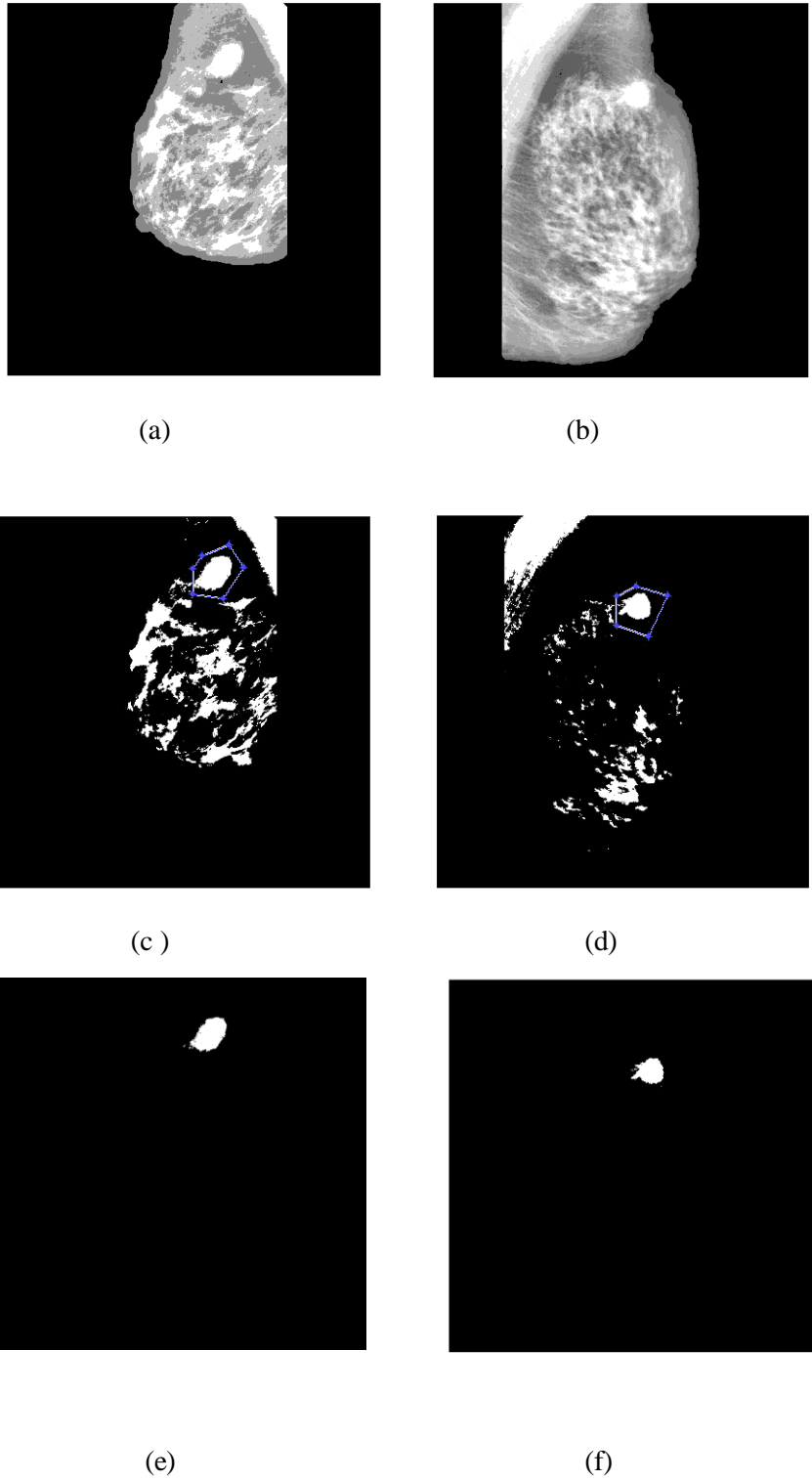
### IV.3. 2. Segmentation

Une fois l'image filtrée et rehaussée, une segmentation est nécessaire pour déterminer sa masse. Nous avons identifié la pathologie ciblée dans notre travail à l'aide de la méthode de k-Means et l'avons ensuite classée comme bénigne ou maligne.

#### ❖ K-Means

La méthode k-Means (voir le chapitre II) renvoie une partition d'une image dans laquelle les individus de chaque cluster sont plus proches les uns des autres et plus loin des individus des autres clusters, chaque cluster de la partition est défini par ces individus et son centre.

Pour cela, nous avons proposé un marquage semi-automatique pour la sélection de la zone d'intérêt comme indiqué en figure IV.8.



**Figure IV.8** - Résultat de la segmentation par k-means des deux images (a) mdb 015 et (b) mdb 202, (c) et (d) sélection de la zone d'intérêt (ROI) et (e) et (f) extraction des ROI.

### IV.3. 3. Descripteurs de forme

L'analyse des images mammographies permet non seulement de détecter les masses mammaires mais aussi de préciser leur nature en matière de malignité ou bénignité en exploitant les caractéristiques extraites de la masse qui sont en général de types texturaux, photométriques et morphologiques. Dans cette étude, on s'est intéressé aux caractéristiques (descripteurs) morphologiques d'aspects géométriques.

À l'issue de la caractérisation, on dispose d'une description unique de chaque structure déterminée par l'algorithme k-Means. Nous avons utilisé 7 descripteurs Parmi lesquels :

- ❖ Area
- ❖ Perimeter
- ❖ centroid
- ❖ Orientation
- ❖ MajorAxislength
- ❖ MinorAxislength
- ❖ Eccentricity

### IV.3.4. Classifieur SVM

Nous avons utilisé un classifieur SVM à deux classes non linéaires (voir le chapitre 03).

L'algorithme nécessite deux étapes:

- ✓ Une étape d'apprentissage dans laquelle le classifieur va apprendre à classer les images de la base d'apprentissage en deux classes (bénigne et maligne) selon le principe cité en chapitre trois.
- ✓ Une étape de reconnaissance dans laquelle le classifieur va prédire la classe à laquelle appartient chaque image de la base de test en utilisant les paramètres trouvés dans la phase précédente.

### IV.3. 4.1. Évaluation des performances de classification

Nous présentons les résultats de la phase de validation dans cette section. Les tableaux de contingence, également appelés matrices de confusion, sont fréquemment utilisés pour évaluer les performances d'un système.

#### ❖ La matrice de confusion

Une matrice de confusion résume les résultats de prédictions d'un problème de classification. En conséquence, les valeurs réelles et les résultats (valeurs prédites) sont comparées. Cela indique non seulement quelles erreurs sont commises, mais surtout le type d'erreur commises (voir tableau IV.1).

		Classe réelle	
		-	+
Classe prédite	-	<b>Vrai Négatif (VN)</b>	<b>Faux Négatif (FN)</b>
	+	<b>Faux Positif (FP)</b>	<b>Vrai Positif (VP)</b>

**Tableau IV.1** - 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.

Plusieurs métriques sont utilisées pour évaluer les performances d'un classifieur. Parmi les quelles, nous citons :

## Chapitre04 : Conception du système d'aide au diagnostic assisté par ordinateur

### ❖ Precision

Le calcul de la précision se fait comme suit :

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

### ❖ Sensibilité

La sensibilité est la proportion de VP sur le nombre total de cas positifs. Elle est calculée comme suit :

$$\text{Sensibilité} = \frac{TP}{TP+FN} * 100$$

### ❖ Spécificité

La spécificité est la proportion de VN sur le nombre total de cas négatifs. Elle est calculée comme suit :

$$\text{Spécificité} = \frac{TN}{TN+FP} * 100$$

## IV.4. Résultats obtenus

Afin d'évaluer les performances du classifieur, nous avons considéré deux cas de figure :

- ❖ **1<sup>er</sup> Cas** : la base d'apprentissage est formée de 80% d'images prises aléatoirement de la base de données et la base de test est constituée de 20% des images restantes.

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

	Bénigne	Maligne
Bénigne	3	0
Maligne	0	5

**Tableau IV.2-** matrice de confusion de notre modèle SVM



## Chapitre04 : Conception du système d'aide au diagnostic assisté par ordinateur

Les performances en termes de précision, de sensibilité et de spécificité sont comme suit :

**Précision= 100% ; Spécificité = 100% et Sensibilité = 100%**

❖ Remarque

Selon les résultats de la matrice de confusion (tableau IV.2), le système a estimé trois masses comme étant bénignes et 5 comme étant malignes. Par conséquent, nous pouvons dire que notre système a parfaitement reconnu les exemples présentés, comme le montre les résultats ci-dessus.

- ❖ **2<sup>ème</sup> Cas** : la base d'apprentissage est formée de 70% d'images prises aléatoirement de la base de données et la base de test est constituée de 30% des images restantes.

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

	Bénigne	Maligne
Bénigne	6	0
Maligne	1	5

Tableau IV.3 : matrice de confusion de notre modèle SVM (2<sup>ème</sup> cas)

Les performances en termes de précision, de sensibilité et de spécificité sont comme suit :

**Précision= 91.6667 % ; Spécificité = 100 % et Sensibilité = 85.71 %.**

## Chapitre04 : Conception du système d'aide au diagnostic assisté par ordinateur

### ❖ Remarque

Les résultats de la matrice de confusion ci-dessus indiquent que le système a estimé 6 masses comme étant bénignes (VN), 5 comme étant malignes (VP) et une masse maligne comme étant bénigne (FP). Le système proposé n'a donc pas bien reconnu tous les exemples présentés, comme le montre bien la valeur de la précision obtenue (91.6667 %). Il est évident que la précision a diminué, et cela est dû au fait que le nombre d'images constituant la base d'apprentissage est moins important dans le deuxième cas (%). Le classifieur n'a pas assez d'exemples pour bien apprendre les classes.

## IV.5. Conclusion

Dans ce chapitre, nous avons mis au point un système d'aide au diagnostic assisté par ordinateur basé sur l'utilisation du k-Means et les machines SVM. L'algorithme proposé donne d'excellents résultats. En effet, afin d'évaluer ses performances, 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 GENERAL**

## CONCLUSION GENERALE

Nous avons mis au point un système semi-automatisé d'aide au diagnostic pour le cancer du sein à partir d'images mammographiques. L'algorithme développé est basé sur l'utilisation d'une segmentation par k-means pour la détection de l'opacité et des Support Vector Machines (SVM) pour la classification de cette dernière comme étant maligne ou bénigne

Dans un premier temps, les images mammographiques ont été prétraitées, afin de supprimer les artefacts et le bruit. Cette opération fondée sur un ensemble de seuillage, amélioration du contraste et d'opérations morphologiques, contribue en particulier à l'amélioration des résultats ultérieurs de segmentation des lésions mammaires.

Ensuite, nous avons utilisé la méthode de k-means pour détecter les opacités de la glande mammaire démontrant ainsi la faisabilité de la segmentation par cette méthode. Après cette opération, viens une phase très importante dans la classification qui permet d'extraire les caractéristiques. On s'est intéressé aux descripteurs morphologiques d'aspects géométriques.

Enfin, nous avons utilisé les machines SVM pour classer la tumeur comme étant bénigne ou maligne. Les résultats obtenus sont très satisfaisants et montrent bien l'efficacité des SVM dans la classification.

Ce mémoire a permis de dégager de nombreuses perspectives que nous pouvons les résumer dans ces quelques lignes :

1. Les attributs morphologiques que nous avons utilisés restent très simples, donc il serait possible de les enrichir par d'autres attributs de plus haut niveau tels que les attributs de texture, de photométrie, etc ...
2. les SVM utilisé permettent juste la classification de deux classes, il est possible d'ajouter une troisième classe concernant le cas « aucune anomalie » qui ne contient pas de tumeur. C'est le type des machines SVM qu'on appelle "multi-classes".
3. On pourrait tester notre système sur une autre base de données telle que la base DDSM.
4. On pourrait également utiliser les réseaux de neurones comme classifieur pour une étude comparative avec les machines SVM.

## **BIBLIOGRAPHIE**

- [01] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-sein>.
- [02] 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.
- [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] 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.
- [06] 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.
- [07] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-sein>
- [08] <https://fr.vecteezy.com/art-vectoriel/6470945-anatomie-du-sein-feminin-vue-de-cote-et-de-face>
- [09] J. Zucman-rossi et J.C.Nault : « Comment une cellule devient cancéreuse et qu'est-ce qu'on peut faire pour l'éviter ». The conversation (academic journalistic flair), 20-12-2016.
- [10] A. Fumleau et D. Mignard sous la coordination de M. Lanta : « Les cancers du sein ». La ligue contre le cancer, Ref : BRK006, 12-2017.
- [11] <https://timm-sante.com/equipements/equipement-mammographie/>
- [12] <https://www.soignez-vous.com/analyses/la-mammographie>

## Bibliographie

---

- [13] S. Chabane et N. Haddache : « Segmentation des images de mammographies ». Mémoire de master, Université Abderrahmane mira Bejaia, faculté des sciences exactes, Département informatique, 2016-2017.
- [14] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>.
- [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] <http://www.breastcare.at/brustkrebsvorsorge/mammographie>.
- [17] B. HAOUZI, « Extraction de réseaux linéiques à partir des images à haute résolution » Memoire en vue de l'obtention du diplome de magister en Telecommunications et Informatique Spatiales 2012.
- [18] D. Ikeda, N. Hylton, H. Kinkel 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.
- [19] J. Bailar, « Mammography : a contrary view. Annals of Internal Medicine ». 84(1) :77–84, (1976).
- [20] K.P Somanand R. Loganathan and V. Ajay. « Machine learning with SVM and other kernel methods ». Eastern Economy Edition, 2009.
- [21] 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 Ille-et- Vilaine ».Edition françaises de radiologie, Paris, vol : 79, pages : 1379-1386, 1998.
- [22] A. Tradivon : « Quels risques, pour quels femmes ?, Densité mammaire et cancer du sein ».Conférence 30es journées de la SFSPM, LA BAULE, 11-2008.
- [23] 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.
- [24] <https://cancer.ca/fr/cancer-information/cancer-types/breast/treatment/surgery>
- [25] <https://hal.science/tel-01762120/document> ( Contributions à l'analyse d'images médicales pour la reconnaissance du cancer du sein) .
- [26] <http://peipa.essex.ac.uk/info/mias.html>.
- [27] D'Osri, C., Bassett, L., Berg, W. et al. « Breast imaging reporting and data system : Acr bi-rads », Technical report, American College of Radiology, (2003).

## Bibliographie

---

- [28] Monsees BS. « The Mammography Quality Standards Act : an overview of the regulations and guidance ». *Radiol Clin North Am* 2000 ; 38 :759–772.
- [29] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh. and Soo, M. S.. Computer-aided detection (cad) in screening mammography :sensitivity of commercial cad systems for detecting architectural distortion. *American Journal of Roentgenology*, 181(4) :1083–1088, (2003).
- [30] <https://perso.esiee.fr/~perretb/I5FM/TAI/morpho/index.html>
- [31] <http://dSPACE.univ-tlemcen.dz/bitstream/112/11262/1/Ms.EBM.Terki%2BBenyelles.pdf>
- [32] [https://fr.wikipedia.org/wiki/Filtre\\_m%C3%A9dian](https://fr.wikipedia.org/wiki/Filtre_m%C3%A9dian)
- [33] R. D. Yapa et Koichi Harada, «Breast Skin-Line Estimation and Breast Segmentation in Mammograms using Fast marching Method,» *International Journal of Biological and Life Sciences*, vol. 3, % 11, pp. 54-62, 2007.
- [34] [https://perso.ensta-paris.fr/~manzaner/Cours/MI206/MI206\\_CH3.pdf](https://perso.ensta-paris.fr/~manzaner/Cours/MI206/MI206_CH3.pdf).
- [35] <https://www.swisscancerscreening.ch/fr/depistage-du-cancer/sein/methodes-de-depistage>.
- [36] <https://memoirepfe.fst-usmba.ac.ma/download/3256/pdf/3256.pdf36>.
- [37] [http://thesis.univ-biskra.dz/864/1/Inf\\_m5\\_2010.PDF](http://thesis.univ-biskra.dz/864/1/Inf_m5_2010.PDF).
- [38]<http://dSPACE.univghardaia.dz:8080/xmlui/bitstream/handle/123456789/1056/memoire%20obenrazek%20hakim.pdf?sequence=1&isAllowed=y>.
- [39] <https://cedric.cnam.fr/vertigo/cours/ml2/docs/coursSVMLLineaires-1x2.pdf>.
- [40] <https://www.memoireonline.com/03/22/12851/Prediction-de-defauts-logiciel-utilisant-des-techniques-dapprentissage-automatique.html>.
- [41] <https://webusers.i3s.unice.fr/~comet/SUPPORTS/Nice-EPU-GB5-IABio/slides-part5.pdf>
- [42] <https://cedric.cnam.fr/vertigo/cours/ml2/docs/coursSVMLLineaires-1x2.pdf>
- [43] <https://zestedesavoir.com/tutoriels/1760/un-peu-de-machine-learning-avec-les-svm/>.
- [44] <https://zestedesavoir.com/tutoriels/1760/un-peu-de-machine-learning-avec-les-svm/>.
- [45] <https://zestedesavoir.com/tutoriels/1760/un-peu-de-machine-learning-avec-les-svm/>.
- [46] <https://zestedesavoir.com/tutoriels/pdf/1760/un-peu-de-machine-learning-avec-les-svm.pdf>.
- [47] <https://zestedesavoir.com/tutoriels/1760/un-peu-de-machine-learning-avec-les-svm/>.
- [48] <https://cedric.cnam.fr/vertigo/cours/ml2/coursMethodesNoyaux.html>
- [49] <https://slideplayer.fr/slide/3834980/>