

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



**Mémoire de Fin d'études Master**

**Filière :** Informatique

**Option :** Systèmes Informatiques

**Thème :**

---

---

**Classification Des Masses Mammographiques Selon  
Leurs Formes**

---

---

**Encadré Par :**

**Dr FERKOUS Chokri**

**Présenté par :**

**DOGHMANE Nabil**

**Juillet 2019**

# Résumé

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

L'objectif de notre travail est de classer les masses mammographiques selon leurs formes (Ronde, Ovale, Lobulaire ou bien Irrégulière). En se basant sur les caractéristiques de forme seulement. La segmentation dans notre approche a été faite manuellement par un radiologue expérimenté à travers un outil que nous avons réalisé qui permet de délimiter les bords de la masse mammographique ; Les contours de la masse sont codés sous forme d'une chaîne de chiffres et enregistrés à la base de données. La phase de description consiste à l'extraction automatique de 10 caractéristiques de formes en se basant sur les contours enregistrés.

Finalement la classification est effectuée à l'aide d'un Perceptron Multicouches (PMC) pour classer les masses segmentées selon leurs formes.

# Abstract

Computer Assisted Diagnostics (DxAO) has become one of the most important research topics in radiology; a typical DxAO system for mammographic classification should include four main steps: pre-treatment, segmentation, feature extraction, and classification.

The objective of our work is to classify the mammographic masses according to their shapes (Round, Oval, Lobular or Irregular). Based on the shape characteristics only. The segmentation in our approach was done manually by an experienced radiologist through a tool that we realized that allows delimiting the edges of the mammographic mass; the contours of the mass are coded as a string of digits and stored in the database. The description phase consists of the automatic extraction of shape characteristics based on the recorded contours.

Finally the classification is carried out using a Perceptron Multilayer (PMC) to classify the masses segmented according to their forms.

# **Remerciement**

*D'abord nous tenons à remercier notre Dieu.  
On profite de l'occasion de la présentation de ce travail  
pour exprimer notre haute gratitude à notre encadreur  
**Mr Chokri FERKOUS***

*qui nous a guidé avec grand patience tout  
au long de ce mémoire et pour son aides précieuses qui ont  
éclairé notre chemin vers le succès et la réussite du travail.  
Merci pour ces conseils, et ca grande gentillesse.*

*J'exprime une réelle reconnaissance aussi à  
**Mr. Beledjhem Nedjmeddine,**  
radiologue à l'hôpital Ibn Rochd d'Annaba (ALGERIE), Pour  
sa disponibilité, ses conseils et ses remarques qui m'ont guidé  
tout au long de ma recherche*

*Nous redressons un grand merci aux membres des jurys qui  
ont bien voulu nous faire l'honneur en acceptant de juger  
ce travail.*

*Nous exprimons enfin nos gratitudes à tous les  
enseignants au niveau du département de l'informatique*

*Nous adressons nos remerciements à ceux qui nous ont  
aidés de près ou loin.*

# Dédicaces

A' mon père Ali  
A' ma mère Halima

A' ma femme Soumeya  
A' ma fille Soundous  
A' mon fils Anas  
A' mon fils Firas  
A' ma fille Tasnim  
A' ma fille Minna

# Table des matières

---

<b>RESUME</b> .....	
<b>ABSTRACT</b> .....	
<b>REMERCIEMENTS</b> .....	
<b>DEDICACE</b> .....	
<b>TABLE DES MATIERES</b> .....	<b>1</b>
<b>LISTE DES FIGURES</b> .....	<b>4</b>
<b>LISTE DES TABLEAUX</b> .....	<b>6</b>
<b>ABREVIATIONS</b> .....	<b>7</b>
<b>INTRODUCTION GENERALE</b> .....	<b>8</b>
<b>CHAPITRE 1: ANALYSE D'IMAGES MAMMOGRAPHIQUES ET DIAGNOSTIQUE DES ANOMALIES MAMMAIRES</b> .....	<b>10</b>
1.1. Introduction.....	10
1.2. Anatomie du sein.....	11
1.3. Le cancer de sein.....	11
1.4. Imagerie du sein .....	12
1.4.1. Échographie du sein .....	12
1.4.2. L'imagerie par résonance magnétique (IRM) .....	13
1.4.3. La mammographie .....	13
1.5. Les anomalies détectables par mammographie.....	16
1.5.1. Les masses.....	16
1.5.1.1. La forme .....	17
1.5.1.2. La densité mammaire.....	17
1.5.1.3. Le contour.....	18

## Table des matières

1.5.2.	Les micro-calcifications .....	18	
1.5.3.	Les distorsions architecturales .....	19	
1.5.4.	L'asymétrie .....	19	
1.6.	La classification des pathologies mammaires .....	20	
1.6.1.	La Classification de Wolfe .....	20	
1.6.2.	La Classification BIRADS .....	21	
1.6.2.1.	BI-RADS 0 .....	21	
1.6.2.2.	BI-RADS 1 .....	21	
1.6.2.3.	BI-RADS 2.....	21	
1.6.2.4.	BI-RADS 3 .....	22	
1.6.2.5.	BI-RADS 4 .....	24	
1.6.2.6.	BI-RADS 5 .....	24	
1.6.2.7.	BI-RADS 6 .....	25	
1.7.	Conclusion.....	26	
<b>CHAPITRE 2: LES CARACTERISTIQUES DE FORME POUR LA CLASSIFICATION</b>			
<b>DES MASSES MAMMOGRAPHIQUES .....</b>			<b>27</b>
2.1.	Introduction.....	27	
2.2.	Les descripteurs de forme en mammographie.....	28	
2.2.1	LES DESCRIPTEURS GEOMETRIQUES (PARAMETRES DE LA FORME) .....	29	
2.2.1.1.	Centre de gravité .....	29	
2.2.1.2.	La circularité.....	32	
2.2.1.3.	Rectangularité .....	32	
2.2.1.4.	La compacité .....	33	
2.2.1.5.	L'aire.....	33	
2.2.1.6.	Le périmètre .....	33	
2.2.1.7.	La méthode de coque convexe.....	34	
2.3.	Processus de classification automatique des masses .....	37	
2.3.1.	La sélection de la région d'intérêt (ROI).....	38	
2.3.2.	La Segmentation des masses mammaires.....	39	
2.3.3.	L'extraction de caractéristiques des masses mammographiques.....	40	
2.3.4.	Taille de la masse .....	43	
2.4.	La classification .....	44	

## Table des matières

2.4.1. Réseaux de neurones .....	44
2.4.2. La classification utilisée .....	48
2.5. La bases De donnée mammographiques .....	49
2.5.1. MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY (MIAS).....	49
2.5.2. DIGITAL DATABASE FOR SCREENING MAMMOGRAPHY (DDSM).....	49
2.5.2.1. DESCRIPTION DU FICHIER .ICS .....	51
2.5.2.2. DESCRIPTION DES FICHIERS ".LJPEG" .....	51
2.5.2.3. DESCRIPTION DU FICHIER OVERLAY .....	52
2.5.2.4. LA DESCRIPTION DES FICHIERS « .16_PGM ».....	52
2.5.2.5. Codage en chaîne.....	53
2.6. Environnement et outils de développement.....	55
2.6.1. Plateformes utilisée .....	55
2.6.2. Langage de programmation et outils .....	55
2.6.2.1. JAVA .....	55
2.6.2.2. Eclipse.....	55
2.7. Réalisation et interface .....	56
2.7. Resultats et discussions .....	57
2.9. Conclusion.....	58
<b>CONCLUSION GENERALE.....</b>	<b>59</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>60</b>



# Liste des Figures

---

Figure 1.1: Anatomie du sein. ....	11
Figure 1.2: Exemple d'image échographique du sein .....	13
Figure 1.3: Exemple d'IRM mammaire .....	13
Figure 1.4: Les composants d'un mammographe. ....	14
Figure 1.5: la masse en mammographie.....	17
Figure 1.6: Les différentes formes d'une opacité mammaire.....	17
Figure 1.7: Différents types de densité mammaire .....	18
Figure 1.8: Les différents types de contours d'une masse mammaire .....	18
Figure 1.9: Exemple d'une distorsion architecturale .....	19
Figure 1.10: La vue des seins droite et gauche .....	19
Figure 1.11: vue Médio Latérale Oblique (MLO).....	20
Figure 1.12: La vue Carnio caudale (CC) .....	20
Figure 1.13: Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2. ....	22
Figure 1.14: Micro-calcifications bénignes : lait calcique : BI-RADS 2.....	22
Figure 1.15: Masses bien circonscrites discrètement lobulées, sans microlobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3. ...	23
Figure 1.16: Microcalcifications granulaires irrégulières peu nombreuses : BIRADS 4.....	24
Figure 1.17: Images spiculée sans centre dense : centre prolifératif d'Aschoff .....	24
Figure 1.18: Masse mal circonscrite aux contours flous et irréguliers : BI-RADS® 5 (cancer canalaire infiltrant SBR II) .....	25
Figure 1.19: Microcalcifications vermiculaires, arborescentes, nombreuses et groupées : BI-RADS 5 .....	24
Figure 2.1: CENTROÏDE DU CONTOUR.....	31
Figure 2.2: ÉCART DE CERCLE.....	31
Figure 2.3: CALCULE DE RECTANGULARITE POUR UNE FORME CIRCULAIRE .....	33
Figure 2.4: CALCULE DE RECTANGULARITE POUR UNE FORME CELLULAIRE.....	33
Figure 2.5: LA COQUE CONVEXE ET LE CONTOUR D'UNE MASSE MAMMAIRE.....	35
Figure 2.6: UN EXEMPLE DE CARACTERISTIQUES A L'AIDE D'UNE MASSE EN ROTATION A DES ANGLES DIFFERENTS .....	37
Figure 2.7: LES ETAPES DE LA METHODOLOGIE PROPOSEE .....	39
Figure 2.8 : LA MAMMOGRAPHIE: A_1103_1.LEFT_CC.LJPEG.1, MASSE MALIGNE, FORME:39	

## Liste des Figures

Figure 2.9: EXEMPLE DE SELECTION DES POINTS $v_i$ .....	40
Figure 2.10: LES FORMES DE MASSES (A) RONDE (B) OVALE (C) LOBULEE (D) IRREGULIERE...	41
FIGURE 2.11: REPRESENTATION D'UN NEURONE ARTIFICIEL.....	43
FIGURE 2.12: SCHEMA D'UN PERCEPTRON MULTICOUCHES (PMC).....	46
FIGURE 2.13: DES ECHANTILLONS DE LA BASE DDSM.....	48
FIGURE 2.14: LES INFORMATIONS DANS UN FICHIER ICS.....	51
FIGURE 2.15: UN EXEMPLE D'UN FICHIER OVERLAY.....	52
FIGURE 2.16: LA VISUALISATION D'UN FICHIER ".16_PGM". .....	53
FIGURE 2.17: CONNECTIVITE 4 VOIES.....	54
FIGURE 2.18: CONNECTIVITE 8 VOIES.....	55
FIGURE 2.19: LA REPRESENTATION DES VALEURS DE DEPLACEMENT DE LA CHAINE CODEE .....	55
FIGURE 2.20: AFFICHAGE DE L'IMAGE ORIGINAL AVEC LA REGION D'INTERET .....	56
FIGURE 2.21: AFFICHAGE DE CONTOUR DE LA PATHOLOGIE.....	57
FIGURE 2.22: AFFICHAGE DE CENTROID DE LA PATHOLOGIE.....	58

# Liste des tableaux

---

Tableau 2.1:.....	42
Tableau 2.2 :.....	44
Tableau 2.3:.....	44
Tableau 2.4:.....	45
Tableau 2.5:.....	52
Tableau 2.6:.....	58
Tableau 2.7:.....	59

# Abréviations

---

<u>ACR</u>	American College of Radiology
<u>CAD</u>	Computed Assisted Diagnosis
<u>BI-RADS</u>	SYSTEME DE COMPTES-RENDUS ET DE DONNEES D'IMAGERIE MAMMAIRE (BREAST IMAGING REPORTING AND DATA SYSTEM)
<u>CC</u>	CRANIO CAUDAL
<u>DA</u>	Distorsions Architecturales
<u>IRM</u>	L'Imagerie par Résonance Magnétique
<u>FN</u>	Faux Négatifs
<u>RR</u>	Risque Relatif
<u>LJPEG</u>	LOSSLESS JPEG
<u>DDSM</u>	LA BASE DE DONNEES NUMERIQUE POUR LA MAMMOGRAPHIE DE DEPISTAGE
<u>DxAO</u>	DIAGNOSTIC ASSISTE PAR ORDINATEUR
<u>MIAS</u>	Mammographic Image Analysis Society
<u>RNA</u>	Réseau de Neurone Artificiel
<u>MCC</u>	MICRO CALCIFICATION
<u>MLO</u>	MEDIO-LATERALE OBLIQUE
<u>PMC</u>	PERCEPTRON MULTICOUCHE
<u>ROI</u>	REGION D'INTERET
<u>SWT</u>	Standard Widget Toolkit
<u>JPA</u>	Java Persistence Api
<u>IDE</u>	Integrated Development Environment

# Introduction Générale

Le développement rapide des technologies d'imagerie médicale est entrain de révolutionner la médecine. L'imagerie médicale permet aux scientifiques e taux médecins d'avoir des informations vitales en examinant le corps humain. L'imagerie médicale joue un rôle de plus en plus important dans le diagnostic et le traitement de la maladie.

Le diagnostic d'images médicales doit être rapide et il doit aussi décrit le plus tôt possible une anomalie surtout dans le cas des cancers. C'est pourquoi notre intérêt s'est porté sur l'extraction des caractéristiques des anomalies, ce qui nous permettra par la suite d'entamer la phase de classification.

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

Selon le système d'aide à la rédaction des comptes rendus BI-RADS [49], les radiologues caractérisent les masses mammographiques en se basant sur leurs densités, leurs formes et leurs types de contours pour les classer dans l'une des classes BI-RADS (2, 3, 4 et 5); ceci est justifié par le fait que les radiologues ne peuvent pas confirmer la malignité d'une anomalie quelle que soit ses caractéristiques sans faire une étude de suivi ou une biopsie.

L'objectif de notre travail est de classifier les masses mammographiques selon leurs formes (Ronde, Ovale, Lobulaire ou bien Irrégulière).En se basant sur les caractéristiques de forme seulement.

Ce mémoire s'articule autour de deux parties qui présentent les différents aspects de ce travail. Dans Le premier chapitre un état de l'art qui présent des généralités du cancer du sein qui nous permettra de comprendre le lexique utilisé dans le domaine, De plus, un aperçu sur les différents aspects pathologiques et radiologiques où nous étudions les différents types d'anomalies selon leurs formes.

## Introduction Générale

Le deuxième chapitre présente une description détaillée de l'approche proposée, ainsi que les résultats expérimentaux obtenus en appliquant notre méthode sur quelques images de la base de données mammographiques DDSM (Digital Database for Screening Mammography).

La segmentation dans notre approche a été faite manuellement par un radiologue expérimenté à travers un outil que nous avons réalisé qui permet de délimiter les bords de la masse mammographique ; Les contours de la masse sont codés sous forme d'une chaîne de chiffres et enregistrés dans la base de données. La phase de description consiste à l'extraction automatique de 10 caractéristiques de formes en se basant sur les contours enregistrés.

Finalement la classification est effectuée à l'aide d'un Perceptron Multicouches (PMC) pour classer les masses segmentées selon leurs formes.

Enfin, Nous terminons notre travail par une conclusion générale qui présente une synthèse des travaux effectués ainsi que les pistes définissant des perspectives possibles pour des travaux futurs.

# Chapitre

# 1

## **Analyse d'images mammographiques et Diagnostique des anomalies mammaires**

### **1.1 Introduction**

Le cancer du sein représente un problème de santé majeur. Il est le type de cancer Le plus fréquent et il se situe en tête de la mortalité chez les femmes dans le monde.Néanmoins, le dépistage précoce de ce cancer augmente le taux de survie et les options de traitement.

La mammographie représente actuellement la meilleure technique utilisée pour le dépistage aussi bien que pour le diagnostic du cancer du sein. En général, ce sont les radiologues en se basant sur leur expérience qui analysent et interprètent les images mammographiques. Néanmoins, malgré leur expertise dans le domaine, des études ont montré qu'ils manquent entre 10% et 30% des cas de cancers du sein [1]. Ceci justifie le besoin de recourir à des systèmes d'aide au diagnostic(DxAO). Ces systèmes guident les radiologues durant le processus de dépistage et accordent une double lecture de la mammographie, ce qui permet une meilleure analyse des cas complexes.

Dans ce chapitre, nous commençons par l'étude de l'anatomie du sein. Cette étude nous permet de comprendre le lexique du domaine et la notion de cancer du sein. Nous abordons par la suite les différentes imageries médicales pouvant être pratiquées sur le sein, en mettant l'accent sur la mammographie comme étant l'imagerie la plus fiable et la plus utilisée ; par la suite nous décrivons les anomalies détectables par la mammographie.

## 1.2 Anatomie du sein

Les seins sont situés au niveau du thorax, en avant du muscle pectoral de chaque côté et s'étendent en hauteur jusqu'à la clavicule et en largeur de l'aisselle jusqu'au milieu du sternum environ. Ils sont asymétriques et ne contiennent pas de muscle. Ils sont soutenus par des ligaments. [2]

Les seins sont constitués d'une glande mammaire, de tissu conjonctif et de tissu graisseux. C'est ce tissu graisseux qui leur donne forme et volume. Recouverts par la peau, ils arborent en leur centre une région pigmentée appelée aréole mammaire qui elle-même entoure le mamelon (orifice d'éjection du lait). La glande mammaire est formée de 15 à 25 lobes qui contiennent des lobules. Ces lobules renferment des unités sécrétrices, les *acini*, qui produisent le lait. Ce dernier est acheminé des lobules jusqu'au mamelon par des canaux sécréteurs, les canaux galactophores [3].

La figure 1.1 détaille la structure du sein et ses différentes composantes.

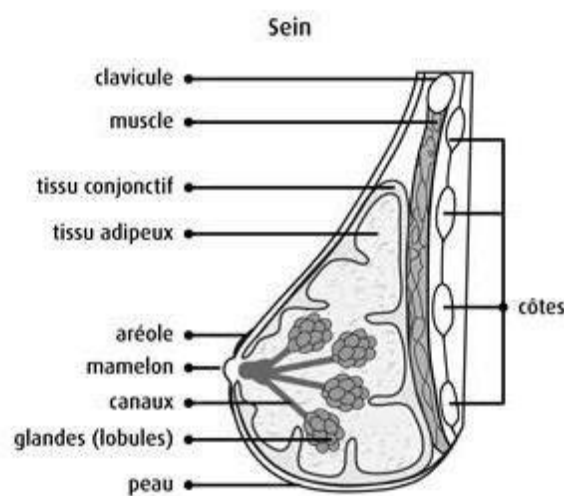


Figure 1.1: Anatomie du sein.

## 1.3 Le cancer du sein

Un cancer du sein résulte d'un dérèglement de certaines cellules qui se multiplient et forment le plus souvent une masse appelée tumeur. Il en existe différents types qui n'évoluent pas de la même manière. Certains sont « agressifs » et évoluent très rapidement, d'autres plus



lentement. Les cellules cancéreuses peuvent rester dans le sein. Elles peuvent aussi se propager dans d'autres organes ce qui est une situation encore plus menaçante. On parle alors de métastases. Dans la majorité des cas, le développement d'un cancer du sein prend plusieurs mois, voire plusieurs années [4].

Le cancer du sein est le résultat d'une multiplication de cellules anormales dans cet organe. Ces cellules sont souvent situées au niveau d'un canal galactophore d'un lobule. Elles peuvent également se retrouver au niveau des ganglions lymphatiques. Les masses formées par cette multiplication excessive sont appelées tumeurs. On peut distinguer deux types de tumeurs :

- **Les tumeurs bénignes** : Ce sont des tumeurs non cancéreuses et ne peuvent se propager ni menacer la vie d'une personne.
- **Les tumeurs malignes** : Contrairement aux tumeurs bénignes, les tumeurs malignes sont des tumeurs cancéreuses. Elles se propagent dans les tissus avoisinants puis dans tout le corps [5]. Cette propagation peut prendre des mois ou même des années [6].

## 1.4 Imagerie du sein

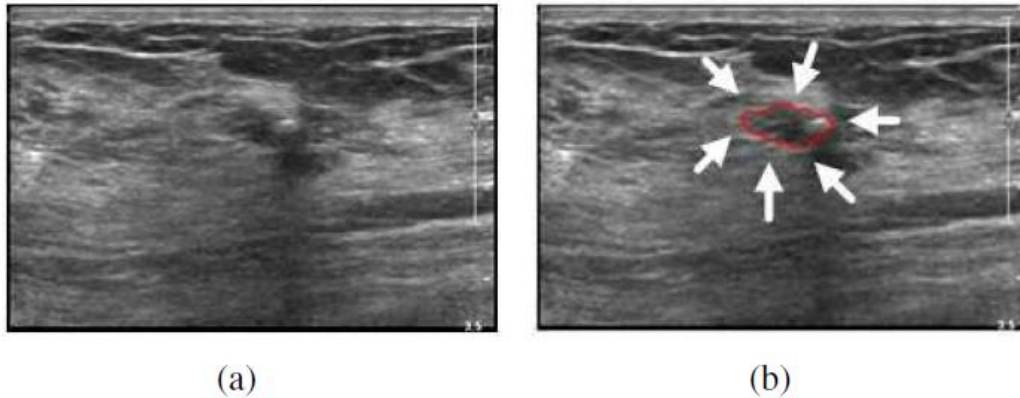
L'imagerie médicale est un domaine de la médecine qui regroupe un ensemble de techniques d'acquisition et de restitution d'images de différents organes du corps humain. Ce domaine a connu d'importants progrès permettant non seulement un meilleur diagnostic, mais offrant aussi de nouvelles possibilités de traitement pour plusieurs maladies.

L'imagerie du sein ou la sénologie radiologique est essentielle pour le dépistage ainsi que pour le diagnostic et la caractérisation d'anomalies visualisées en examen de dépistage. La mammographie est l'examen paraclinique de référence en sénologie, éventuellement couplée à une échographie mammaire. Parfois, une étude complémentaire est nécessaire. Ainsi, un bilan IRM (Imagerie par Résonance Magnétique) peut être indiqué [7]. Nous présentons dans ce qui suit ces différents outils.

Trois types d'imagerie médicale sont utilisés pour le sein : l'échographie, l'imagerie par Résonance Magnétique (IRM) et la mammographie.

### 1.4.1 Échographie du sein

Une échographie mammaire est une technique d'imagerie couramment utilisée pour dépister des tumeurs et d'autres anomalies mammaires. L'échographie utilise des ondes sonores à haute fréquence pour produire des images détaillées de l'intérieur des seins. Contrairement aux rayons X et à la tomodensitométrie, les ultrasons n'utilisent pas la radiation et sont considérés comme sûrs pour les femmes enceintes et les mères qui allaitent.

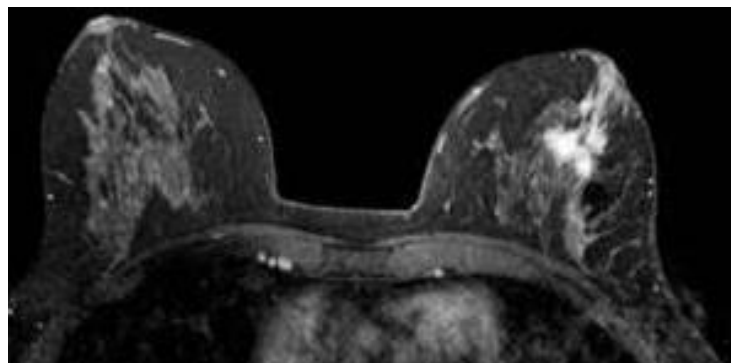


*Figure 1.2 : Exemple d'image échographique du sein.*

L'exemple de la détection et de la segmentation d'une tumeur du sein. (a) L'image originale avec une tumeur bénigne, (b) Segmentation manuelle de la tumeur (ligne rouge).

#### **1.4.2 L'imagerie par résonance magnétique (IRM)**

L'imagerie par résonance magnétique (IRM) est une méthode permettant d'obtenir des images de l'intérieur d'objets, en particulier d'objets vivants tels que l'homme et les animaux. Il n'utilise pas de radiations ionisantes telles que les rayons X. Au lieu de cela, il utilise des ondes radioélectriques (RF) et un champ magnétique intense pour exciter les atomes de l'objet en cours d'évaluation. Les motifs de cette excitation sont observés sur un écran. L'IRM peut fournir des vues tridimensionnelles en temps réel des organes du corps, des muscles et des articulations sans chirurgie invasive [13].



*Figure 1.3 : Exemple d'IRM mammaire*

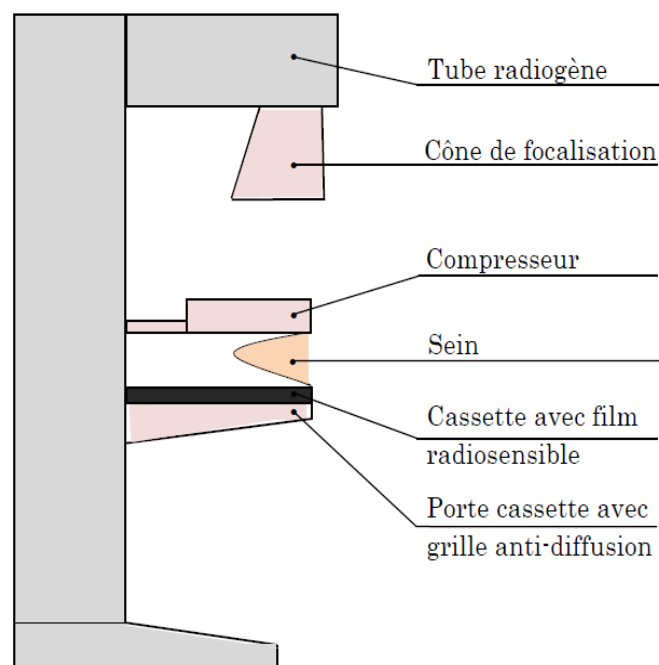
#### **1.4.3 La mammographie**

La mammographie est une méthode d'imagerie par rayons X utilisée pour examiner le sein afin de détecter rapidement le cancer et d'autres maladies du sein. Il est utilisé à la fois comme outil de diagnostic et de dépistage.

Elle permet de détecter la présence d'anomalies pouvant être le signe d'un cancer du sein [14]. Cet examen est réalisé en utilisant une mammographie : il s'agit d'un appareil formé par un générateur de Rayons X de faible énergie et d'un système de compression du sein permettant l'étalement des tissus mammaires ce qui facilite la visualisation des structures du sein et permet de réduire la dose de rayons X délivrée.

Pendant une mammographie, le sein du patient est placé sur une plaque de support plate et comprimé avec une plaque parallèle appelée palette. Un appareil à rayons X produit une petite rafale de rayons X qui traverse le sein vers un détecteur situé du côté opposé. Le détecteur peut être soit une plaque de film photographique, qui capture l'image de rayons X sur un film, soit un détecteur à l'état solide, qui transmet des signaux électroniques à un ordinateur pour former une image numérique. Les images produites s'appellent des mammographies.

Sur une mammographie de film, les tissus à faible densité, tels que les tissus adipeux, apparaissent translucides (c.-à-d. Des nuances de gris plus foncées se rapprochant du fond noir), tandis que les zones de tissu dense, telles que les tissus conjonctifs et glandulaires ou les tumeurs, apparaissent plus blanches sur un fond gris. Dans une mammographie standard, on prend à la fois une vue de dessus et une vue de côté de chaque sein, bien que des vues supplémentaires puissent être prises si le médecin est préoccupé par une zone suspecte du sein.



**Figure 1.4 :** Les composants d'un mammographe

La mammographie reste aujourd'hui le meilleur examen pour le dépistage précoce du cancer du sein. Cependant, cet examen comporte des limites, pouvant être liées à de réelles

difficultés d'interprétation (dus à la tumeur ou à la structure mammaire environnante) ou à des facteurs liés au lecteur, comme la fatigue, la distraction ou le niveau d'expérience du radiologue (15). Les lésions manquées sont souvent liées à une insuffisance de détection du radiologue.

Ceci concerne tout particulièrement la détection des foyers de micro-calcifications qui peuvent ne pas être vus lorsque la vision n'est pas focalisée sur l'anomalie (16).

Le taux de faux négatifs en première lecture est loin d'être négligeable puisqu'il a été évalué entre 16 et 31 %. Afin de réduire ce chiffre, une relecture systématique des mammographies de dépistage est pratiquée par des radiologues validant le cahier des charges. Cette double lecture réduit le taux de faux négatifs de 5 à 15 %, en améliorant particulièrement la détection des cancers de petite taille (17). Le bénéfice est d'autant plus grand que les deuxièmes lecteurs sont plus spécialisés, lisent plus de clichés et ont reçu une formation spécifique. La double lecture est une procédure acquise pour les examens mammographiques de fessant de radiologues spécialisés dans la lecture des mammographies (18). Le concept de systèmes experts informatiques dédiés à l'aide au diagnostic (CAD = Computed Assisted Diagnosis)

La mammographie, pratiquée régulièrement et dans le cadre des programmes de dépistage, permet de diagnostiquer la maladie à un stade suffisamment précoce pour que le traitement soit le plus conservateur possible et en même temps efficace. En effet tant que la taille de la tumeur est inférieure à 1-2 cm, les chances de guérison sont voisines de 100 % [8].

Le dépistage par mammographie est l'un des moyens les plus efficaces de dépistage du ont souvent utilisé des outils de diagnostic. La mammographie, cependant, n'est pas efficace pour les seins denses, avec des taux élevés de faux négatifs [9]. Au cours des dernières années, l'imagerie ultrasonore est devenue un outil important pour la détection précoce des tumeurs mammaires, qui prouve des informations diagnostiques complémentaires pour la mammographie, en particulier chez les patientes dont les tissus mammaires sont denses [10].

Pour évaluer des tumeurs du sein par échographie, les radiologues considèrent plusieurs les caractéristiques de l'image, telles que la forme de la lésion, l'orientation, et l'amélioration acoustique postérieure [11].

L'interprétation des images échographiques est toutefois subjective. Et la variabilité est très élevée en raison de sa faible résolution d'image et les différentes expériences des radiologistes qui analysent les données des caractéristiques de la tumeur. Variations dans la perception humaine de l'image différentes caractéristiques utilisées dans l'interprétation de l'image, et

l'absence de mesures quantitatives des caractéristiques utilisées pour l'analyse d'images, entraînent une variabilité diagnostique entre radiologues. Par conséquent, la confirmation finale est obtenue par d'autres modalités, comme l'IRM ou la biopsie [12].

- **Faux négatif**

Un « *faux négatif* » signifie que les résultats de la mammographie n'ont pas détecté la présence d'un cancer alors qu'il y en a bien un. Les femmes concernées ne savent pas qu'elles ont un cancer, d'où l'importance que l'examen clinique des seins soit fait par le médecin. Un cancer du sein peut ne pas être détecté si la tumeur est très petite, si elle est localisée dans un endroit difficile à visualiser par mammographie (comme l'aisselle) ou si elle est cachée par d'autres images.

- **Faux positif**

Un « *faux positif* » signifie que les résultats de la mammographie sont interprétés comme anormaux alors qu'en réalité il n'y a pas de cancer.

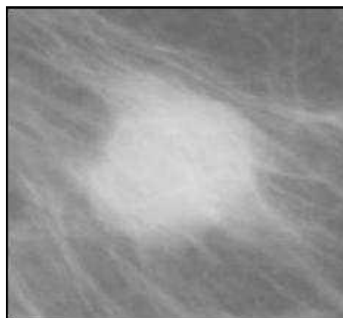
- **Vrai Positif** : est le nombre de lésions malignes qui sont classées malignes.
- **Vrai Négatif** : est le nombre de lésions bénignes qui sont classées bénignes.

## 1.5 Les anomalies détectables par mammographie

Les différents types de cancer du sein ont été décrits par l'ACR (American College of Radiology) [19-20]. Parmi les catégories d'anomalies repérables sur les images de mammographie nous pouvons citer les masses, les micro-calcifications, les distorsions architecturales et l'asymétrie.

### 1.5.1 Les masses :

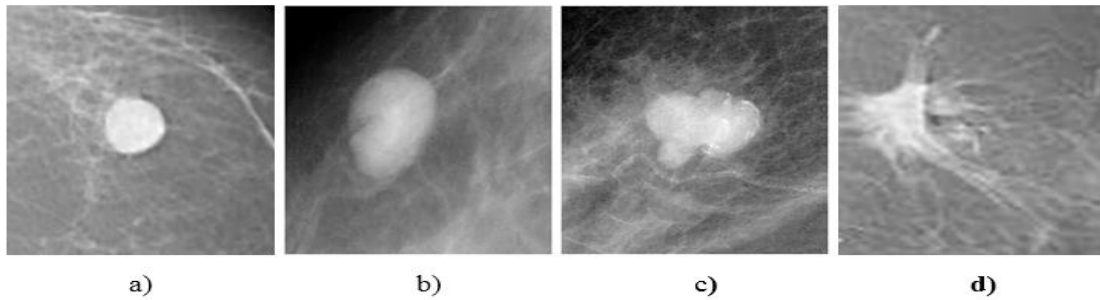
Une masse est une lésion occupant un espace sur le sein. Sa détection est considérée comme une tâche difficile [21]. La caractérisation d'une masse est basée sur sa forme (ronde, ovale, lobulée, irrégulière), sur ses contours (circonscrits, micro-lobulés, masqués, indistincts, spicules) et sur sa densité (haute, moyenne, faible, grasseuse) [22].



*Figure 1.5 : la masse en mammographie*

### 1.5.1.1 La forme:

Quatre formes différentes peuvent être distinguées : ronde, ovale, lobulaire et irrégulière, (Figure 1.6).



*Figure 1.6 : Les différentes formes d'une opacité mammaire*

- a) Ronde : Il s'agit de masse sphérique, circulaire ou globuleuse.
- b) Ovale : Elle présente une forme elliptique
- c) Lobulée : La forme de la masse présente une légère ondulation.
- d) Irrégulière : Cette appellation est réservée aux masses dont la forme est

### 1.5.1.2 La densité mammaire:

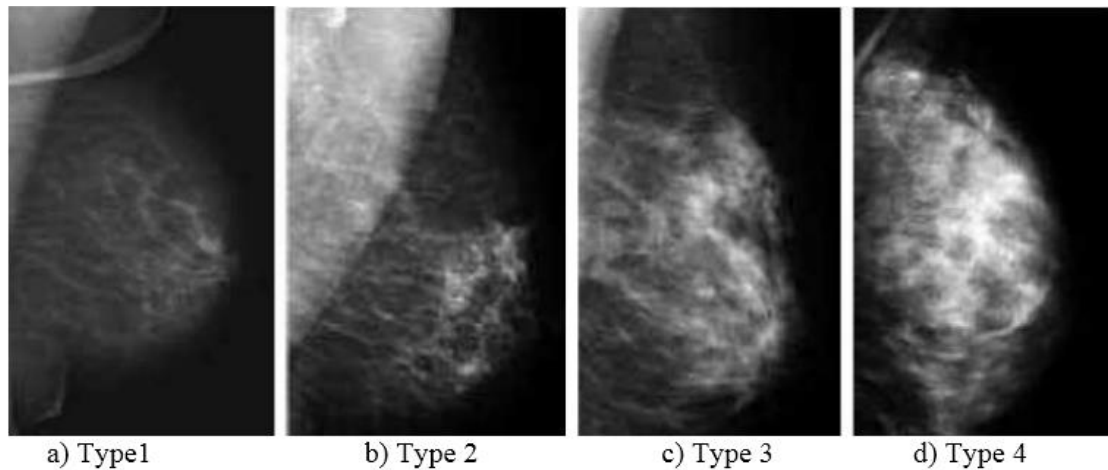
Les masses sont généralement plus denses par rapport au tissu adjacent. La visibilité d'une masse varie en fonction de la densité du sein.

Le sein est constitué de différents types de tissus : le tissu glandulaire et conjonctif et la graisse. La densité des seins décrit la quantité relative des différents tissus dans le sein. Les seins denses présentent moins de tissu graisseux et plus de tissu conjonctif et glandulaire.

La recherche montre que le risque de présenter un cancer du sein est 4 à 6 fois plus élevé chez les femmes qui ont des seins denses que chez celles qui n'ont pas des seins denses [23].

Le tissu mammaire change au fur et à mesure que nous vieillissons ; il devient généralement moins dense avec l'âge, bien que certaines femmes aient toujours des seins denses sans égard à leur âge [24].

La classification BI-RADS( Breast Imaging Reporting and Data System ) de L'American College of Radiology permet de classer les seins en quatre groupes pour lesquelles le risque de rater un cancer en raison du caractère hétérogène du tissu fibro-glandulaire augmente lorsque l'on passe de la catégorie 1 à la catégorie 4 (figure suivante) [25]:

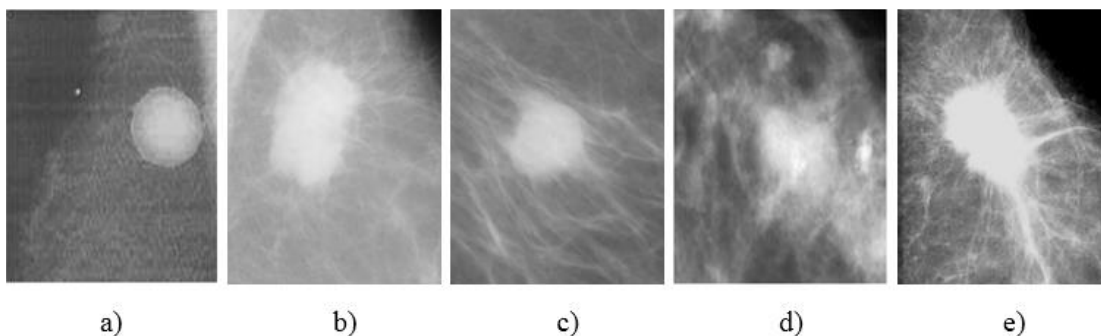


**Figure 1.7 :** *Différents types de densité mammaire. [26]*

Quelle que soit la méthodologie utilisée pour évaluer la densité mammaire, toutes les études publiées concernant la population générale (plus grande hétérogénéité des données dans les populations symptomatiques) ont retrouvé un sur risque de cancer du sein chez les femmes ayant des seins denses [28]. Une méta-analyse regroupant 42 études [27] retrouve un risque relatif (RR) de 1,79 pour une densité de 5 à 24 %, de 2,11 pour une densité de 25 à 49 %, de 2,92 pour une densité de 50 à 74 %, et de 4,64 pour une densité supérieure à 75 % (tableau). Avec la classification de Wolfe, les études incidentes (mammographie négative avant le diagnostic de cancer) et prévalentes (analyse du sein controlatéral lors du diagnostic du cancer du sein)

### 1.5.1.3 Le contour :

Cinq types de contour peuvent caractériser une masse : circonscrit, effacé, micro-lobulé, mal-défini et spiculé. (Figure 1.8).



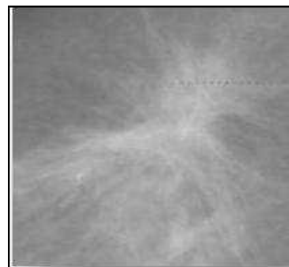
**Figure 1.8:** *Les différents types de contours d'une masse mammaire [29]*

### 1.5.2 Les micro-calcifications

En général, ce sont des dépôts de sels de calcium, elles apparaissent sous forme de points blancs et brillants sur les mammogrammes. Les microcalcifications sont considérées comme premiers signes de malignité, la variation de forme et la petite dimension qui caractérise une micro-calcification complique la tâche de sa détection [30].

### 1.5.3 Les distorsions architecturales

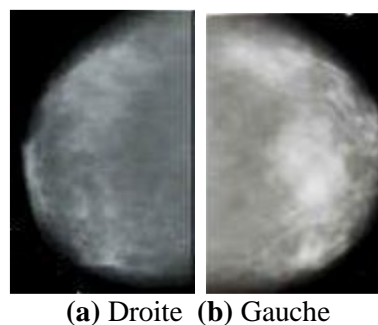
Une distorsion architecturale est un changement au niveau de l'architecture normale du tissu mammaire, Souvent, ces DA peuvent être présentées avec les masses ou les microcalcifications, mais elles sont moins fréquentes que ces deux dernières. Il est estimé que dans les dépistages de mammographie entre 12 % et 45% des cancers ne sont pas détectés à cause des DA [31-32]. Un exemple de distorsion architecturale est donné dans la figure (2.8). [33]



*Figure 1.9: Exemple d'une distorsion architecturale*

### 1.5.4 L'asymétrie

L'asymétrie de densité est définie par une différence de densité entre les seins. Cette différence doit être observée sur les différents clichés aussi bien en incidence oblique (MLO) qu'en incidence de face (CC) [34]. Cette asymétrie est, généralement, à l'origine de l'existence d'une masse [35].



*Figure 1.10 : La vue des seins droite et gauche*

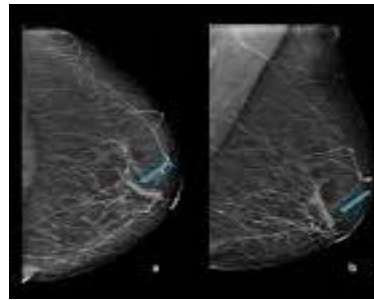


La vue Médio Latérale Oblique (MLO) : son acquisition est réalisée avec une position oblique par rapport au sein. Lors de la capture de ce type de vue, une région appartenant au muscle pectoral peut être présente dans la mammographie. Cette région se trouve au coin supérieur gauche ou au coin supérieur droit de l'image (figure 1.11).



**Figure 1.11** : La vue Médio Latérale Oblique (MLO).

La vue Carnio caudale (CC) : la position du sein lors de l'acquisition de la vue CC doit être horizontale (figure 1.12). Notons que cette vue montre autant que possible les tissus glandulaires (les canaux et les lobules) ainsi que les tissus gras qui entourent les tissus glandulaires.



**Figure 1.12** : La vue Carnio caudale (CC).

## **1.6 La classification des pathologies mammaires**

### **1.6.1 Classification de Wolfe**

Wolfe le premier en 1976, essayé d'établir un lien entre l'aspect du sein et le risque de cancer [36]. Il parvint à classer les images mammographiques en quatre classes en fonction de la densité mammaire et de la répartition du tissu conjonctif :

N1 : l'aire mammaire est conjonctif avec très peu de tissus denses

P1 : les éléments denses occupent moins du quart de l'aire mammaire

P2 : les éléments denses occupent plus du quart de l'aire mammaire

DY : l'aire mammaire est globalement dense

Selon A. Wolfe, seules les classes P2 et DY présentent une forte susceptibilité de pathologie.

## **1.6.2 Classification BIRADS**

BI-RADS est un système d'aide à la rédaction des comptes rendus de plus en plus utilisé dans le monde et bientôt implanté directement sur les matériels de mammographie et d'échographie. Les catégories d'évaluation du BI-RADS® permettent une synthèse claire des données descriptives issues de l'utilisation du lexique et invitent le radiologue à une démarche raisonnée, objective et moins intuitive. Elles donnent une conduite à tenir et responsabilisent le radiologue et les autres intervenants dans la prise en charge des patientes.

Selon la norme adoptée par le système BIRADS, il existe sept types de tissus mammaires.

### **1.6.2 .1 BI-RADS 0**

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

### **1.6.2 .2 BI-RADS 1**

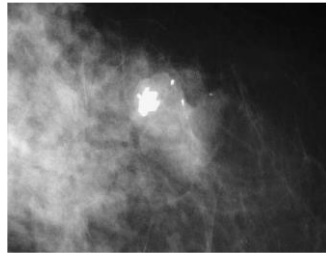
L'examen est négatif : mammographie normale. Les seins sont symétriques et il n'y a pas d'opacité, de distorsion architecturale ou de calcification suspecte. En principe cela ne pose aucun problème si les seins sont radio-transparents. Mais quelle est la certitude en cas de seins denses (type 3 ou 4) ? Les solutions possibles résident dans la comparaison avec les mammographies anciennes, l'examen clinique – ce qui renvoie à la question plus générale de l'intérêt de l'échographie systématiquement associée à l'exploration des seins denses.

### **1.6.2 .3 BI-RADS 2**

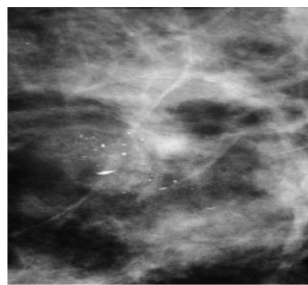
Constatations bénignes : « Cela correspond pratiquement à une mammographie négative mais le radiologue désire décrire une anomalie dont les caractères radiologiques sont caractéristiques et qui peuvent être étiquetées avec beaucoup de confiance dans l'interprétation. » Le risque d'association à un cancer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

- masses rondes avec macrocalcifications (fibroadénome / kystes) (Figure 1.13) ;
- masses rondes ou ovales à centre clair (ganglion intramammaire);
- masse ronde correspondant à un kyste typique en échographie;

- anomalie de densité grasseuse ou mixte (kyste huileux, lipome, galactocèle, hamartome);
- cicatrice connue et calcification sur matériel de suture ;
- macrocalcifications sans opacité (adénofibrome, kyste, adiponécrose, ectasie canalaire sécrétante, calcifications vasculaires etc.) ;
- microcalcifications annulaires ou arciformes, semilunaires, sédimentées, rhomboédriques (Figure 1.14) ;
- calcifications cutanées et calcifications punctiformes régulières diffuses.



*Figure 1.13 : Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2.*

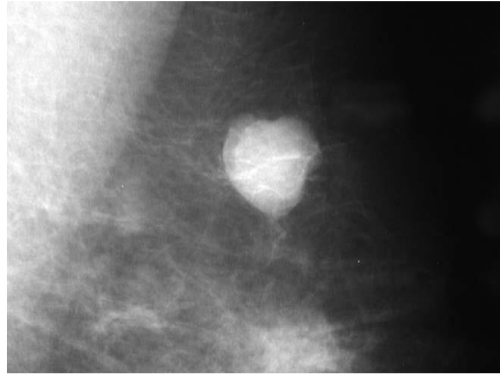


*Figure 1.14 : Micro-calcifications bénignes : lait calcique : BI-RADS 2.*

#### **1.6.2 .4 BI-RADS 3**

Anomalie probablement bénigne (> 98 %). Proposition d'une surveillance à court terme. Il s'agit de la catégorie la plus délicate à manier du fait de l'incertitude qu'elle comporte.

Il n'y a pas de modification attendue pendant la période de surveillance et les données de la littérature sont en faveur de cette attitude. La surveillance mammographique est justifiée par le fait que ces lésions probablement bénignes ont effectivement un très faible risque d'être malignes; que la mammographie identifiera, parmi les rares lésions qui se modifient pendant l'intervalle, celles qui sont effectivement malignes ; que ces cancers seront diagnostiqués tôt dans leur évolution à un stade encore favorable.



*Figure 1.15 : Masses bien circonscrites discrètement lobulées, sans microlobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3.*

Ces lésions doivent être non palpables et ne pas être nouvelles ou en progression par comparaison avec une mammographie antérieure. Il faut préciser dans le compte rendu le rythme suggéré de surveillance. Dans le cahier des charges du dépistage français, il est recommandé un contrôle entre quatre et six mois deux fois puis tous les ans pendant au moins deux ans. Au terme d'une surveillance suffisante pendant deux à trois ans, la lésion stable peut être reclassée BI-RADS® 2. En cas de modification au cours de la surveillance, la biopsie est la règle. Les données de la littérature sont en faveur de cette attitude puisque le risque de cancer observé dans la littérature est de 0,3 à 1,7 % des lésions classées BI-RADS® 3. Mais il arrive parfois qu'une biopsie soit pratiquée d'emblée notamment si une grossesse est désirée ou en cours, si un traitement hormonal substitutif est envisagé, s'il existe des facteurs de risque familiaux ou personnels ou un risque de mauvaise compliance à la surveillance [37]. L'intérêt de l'échographie devant une masse probablement bénigne est évident : quand elle permet de faire le diagnostic immédiat de kyste ou à l'opposé d'une lésion hautement suspecte, elle évite respectivement une surveillance inutile ou une attente délétère de six mois.

Les anomalies susceptibles d'être classées BIRADS 3 sont :

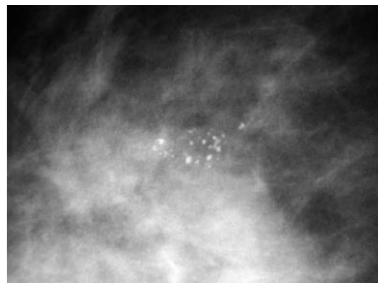
- masses ou opacités bien circonscrites rondes ou ovales ou discrètement lobulées, sans microlobulation, non calcifiées, non typiquement liquidiennes en échographie (Figure. 1.15) ;
- asymétries focales de densité à limites concaves et/ou mélangées à de la graisse ;
- petits amas ronds ou ovales de calcifications amorphes, peu nombreuses, évoquant un début de calcification d'adénofibrome ;
- une discrète distorsion sans opacité centrale dans la zone d'une biopsie antérieure;

- microcalcifications rondes ou punctiformes régulières ou amorphes, peu nombreuses, en petit amas rond isolé.

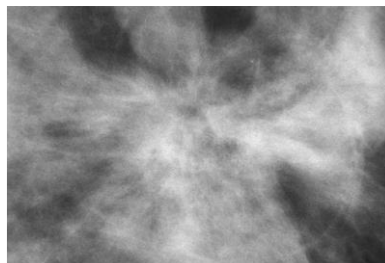
#### 1.6.2 .5 BI-RADS 4

Anomalie suspecte : une biopsie doit être envisagée. Ce sont des lésions qui n'ont pas toutes les caractères morphologiques d'un cancer typique mais qui ont une probabilité importante d'être malignes [38]. Cette probabilité est de 10 à 50 % [39]. Une vérification histologique est nécessaire. Ce sont les :

- microcalcifications punctiformes régulières nombreuses et/ou groupées en amas aux contours ni ronds ni ovales ;
- microcalcifications amorphes, groupées et nombreuses ;
- microcalcifications irrégulières, polymorphes ou granulaires, peu nombreuses (Figure1.16)
- distorsion architecturale sans centre dense (Figure1.17) en dehors d'une cicatrice connue et stable ;
- masses opaques solides (échographie) rondes ou ovales à contours lobulés ou masqués ou ayant augmenté de volume ;
- asymétries de densité focales, à limites convexes ou évolutives.



**Figure1.16** :Microcalcifications granulaires irrégulières peu nombreuses : BIRADS 4.

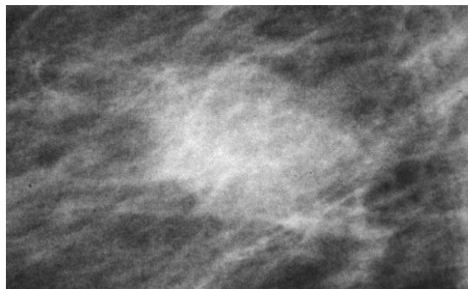


**Figure1.17** : Images spiculée sans centre dense.

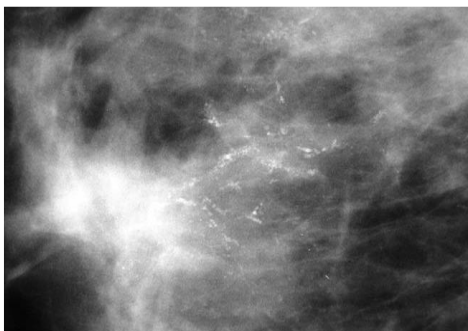
#### 1.6.2.6 BI-RADS 5

Haute probabilité de malignité : une action appropriée doit être entreprise. Ces lésions ont une très forte probabilité de correspondre à un cancer du sein (> 95 %) [40]. Il s'agit de :

- masses spiculées à centre dense ;
- masses irrégulières à contours indistincts ou spiculés (Fig. 1.18) ;
- microcalcifications vermiculaires, arborescentes ou microcalcifications irrégulières, polymorphes ou granulaires, nombreuses et groupées (Fig. 1.19) ;
- groupements de microcalcifications quelle que soit leur morphologie, dont la topographie est galactophorique ;
- microcalcifications associées à une distorsion architecturale ou à une opacité ;



**Figure 1.18 :** Masse mal circonscrite aux contours flous et irréguliers : BI-RADS® 5 (cancer canalaire infiltrant SBR II).



**Figure 1.19 :** Microcalcifications vermiculaires, arborescentes, nombreuses et groupées : BI-RADS 5.

- microcalcifications groupées ayant augmenté en nombre ou microcalcifications dont la morphologie et la distribution sont devenues plus suspectes.

Une biopsie guidée par l'imagerie est utile pour confirmer le diagnostic, d'autant plus que le traitement envisagé comporte la technique du ganglion sentinelle ou une chimiothérapie adjuvante

### 1.6.2.7 BI-RADS 6

Résultat de biopsie connu c'est à dire malignité prouvée.

## **1.7 Conclusion**

Dans ce chapitre, nous avons introduit quelques notions générales concernant l'anatomie et le cancer de sein ainsi que les outils de l'imagerie médicale permettant le diagnostic de ce type de cancer ont été présentés. Une grande attention a été consacrée à l'étude des spécifications des pathologies mammaires à savoir les masses.

Une telle étude est fortement associée à la présentation des standards adoptés par les radiologues pour classifier les lésions mammaires spécialement selon la norme BI-RADS. L'objectif de détailler ces études est de comprendre le lexique utilisé par les radiologues et qui nous permettra par la suite de choisir les méthodes d'extraction de caractéristiques adéquates pour le diagnostic assisté par ordinateur (DxAO). Nous détaillons dans le chapitre suivant les étapes de notre système DxAO.

# Chapitre

# 2

## Les caractéristiques de forme pour la classification des masses mammographiques

### 2.1 Introduction

La forme d'un objet est une caractéristique visuelle importante et fondamentale pour décrire le contenu de l'image, et peut être considérée comme une silhouette de l'objet [41], ce descripteur doit être invariable à la rotation, l'échelle et la translation[42]. Les formes sont souvent les archipels d'objets appartenant à la même classe de paternité, et peuvent être utilisées dans un large éventail de problèmes pratiques, tels que l'analyse de documents (reconnaissance optique de caractères), les arts visuels (restauration vidéo), Internet (récupération d'images à partir du contenu), la sécurité (détection des empreintes digitales), etc. [43]. Dans les systèmes d'extraction d'images basés sur le contenu, par exemple, les descripteurs de forme peuvent être utilisés pour coder les propriétés d'image qui sont pertinentes à une requête. Pour les tâches d'enregistrement d'images, on peut utiliser la forme pour identifier les correspondances entre les contours des objets.

Le rôle d'interprétation des images médicales est de fournir au médecin des Informations symboliques de haut niveau sur le contenu de l'image. Dans un but d'aide au diagnostic, ces informations doivent expliciter la région d'intérêt (ROI) ainsi que les aspects pathologiques des structures présentes dans l'image.

L'interprétation des images médicales représente un outil efficace pour le traitement de l'information visuelle, elle permet inévitablement non seulement de détecter et de localiser



la tumeur mais aussi de spécifier ou plutôt de prévoir la gravité de la tumeur, en terme de malignité/bénignité, en exploitant les caractéristiques extraites de l'image.

L'interprétation d'images est un problème complexe qui peut être défini comme l'extraction automatique de la sémantique d'une image. Cependant, cette sémantique n'est pas toujours explicitement dans l'image elle-même. Elle dépend d'une part des connaissances a priori sur le domaine et d'autre part du contexte de l'interprétation.

Le travail présenté dans ce chapitre vise l'extraction des caractéristiques et la classification des masses mammographiques selon leurs formes (Ronde, Ovale, Lobulaire ou bien Irrégulière).

## 2.2 Les descripteurs de forme

Les caractéristiques représentent des transformations mathématiques qui reformulent l'aspect pathologique des masses pour pouvoir être utilisées par l'algorithme de classification. Par conséquent, l'extraction de caractéristiques affecte les résultats de classification vu que plus la caractérisation des masses est bonne plus le résultat de classification sera précis. La description et la modélisation de la forme des masses peuvent être faites de deux façons ; Soit en se basant sur le contour soit en utilisant l'intérieur de la masse. Néanmoins, Le contour peut être obtenu à partir de l'intérieur et vice versa (l'intérieur peut être Reconstitué à partir du contour). Par conséquent, les deux méthodes de modélisation de Forme sont équivalentes.

Les caractéristiques de forme efficaces doivent présenter certaines propriétés essentielles telles que :

- **Identifiabilité:** Les formes qui se trouvent perceptuellement similaires par l'homme ont la même caractéristique différente des autres.
- **Translation :** Est l'invariance de l'emplacement, de la rotation et du changement d'échelle de la forme, cette propriété ne doit pas affecter les caractéristiques extraites.
- **Invariance affine :** la transformation affine effectue une application linéaire des coordonnées 2D vers d'autres coordonnées 2D qui préserve la "rectitude" et le "parallélisme" des lignes. La transformation affine peut être construite en utilisant des séquences de translations, d'échelles et de rotations. Les caractéristiques extraites doivent être aussi invariantes que possible avec la transformation affine.

- **Résistance au bruit** : les fonctions doivent être aussi robustes que possible contre le bruit, c'est-à-dire qu'elles doivent être les mêmes quel que soit la force du bruit dans une plage donnée qui affecte le motif.
- **invariance à l'occultation** : lorsque certaines parties d'une forme sont occultées par d'autres objets, la caractéristique de la partie restante ne doit pas changer par rapport à la forme originale.
- **Statistiquement indépendant**: deux caractéristiques doivent être statistiquement indépendantes. Ceci représente la compacité de la représentation.
- **Fiable**: tant que l'on traite le même motif, les caractéristiques extraites doivent rester les mêmes.

En général, un descripteur de forme est un ensemble de valeurs produites pour décrire une caractéristique de forme donnée. Un descripteur tente de quantifier la forme d'une manière qui soit en accord avec l'intuition humaine (ou avec des exigences spécifiques à une tâche).

La précision nécessite un descripteur de forme pour pouvoir trouver efficacement des formes similaires dans une base de données. Habituellement Les descripteurs sont sous la forme d'un vecteur.

### **2.2.1 Les descripteurs géométriques (Paramètres de la forme)**

Fondamentalement, la récupération d'images basée sur la forme consiste à mesurer la similarité entre les formes représentées par leurs caractéristiques.

Certaines caractéristiques géométriques simples peuvent être utilisées pour décrire des formes. Habituellement, les caractéristiques géométriques simples peuvent seulement distinguer les formes avec de grandes différences; par conséquent, ils sont généralement utilisés comme filtres pour éliminer les fausses classifications de forme ou les combinaisons avec d'autres descripteurs de forme pour distinguer les formes. Ils ne sont pas appropriés pour être des descripteurs de forme autonomes. Une forme peut être décrite par différents aspects.

#### **2.2.1.1 Centre de gravité :**

Le centre de gravité s'appelle également centroïde. Sa position devrait être fixée par rapport à la forme et souvent utilisé pour définir les descripteurs de forme, ainsi que pour

Les caractéristiques de forme pour la classification des masses mammographiques

définir d'autres descripteurs géométriques tels que la distance radiale, Si une forme est représentée par sa fonction de région, son centroïde ( $g_x$ ;  $g_y$ ) est:

- Représentation de la forme par la région de l'objet en sa totalité :

$$g_x = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{Équation 2.1}$$

$$g_y = \frac{1}{N} \sum_{i=1}^N y_i \quad \text{Équation 2.2}$$

Où N est le nombre de points de la forme,  $(x_i, y_i) \in \{(x_i, y_i) | f(x_i, y_i) = 1\}$ .

Si une forme est représentée par son contour, La position de son centroïde est donnée ci-dessous:

- Représentation de la forme par le contour de l'objet :

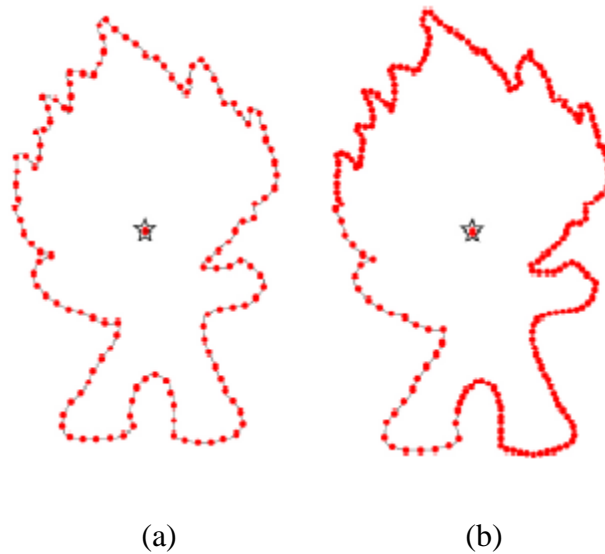
$$g_x = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad \text{Équation 2.3}$$

$$g_y = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \quad \text{Équation 2.4}$$

- Avec A est la surface du contour et

$$A = \frac{1}{2} \sum_{i=0}^{N-1} (x_i y_{i+1} - x_{i+1} y_i) \quad \text{Équation 2.5}$$

La position du centre de la forme est fixée par la distribution de points de contours. Nous pouvons remarquer que la position du centroïde sur la figure1 est fixe quelle que soit la répartition des points autour de la forme.



**Figure 2.1 :** Centroïde du contour

Les points sont répartis sur le contour de manière uniforme (a) et non uniforme (b). L'étoile est le centroïde des contours; le point intérieur est le centre de gravité du contour échantillonné.

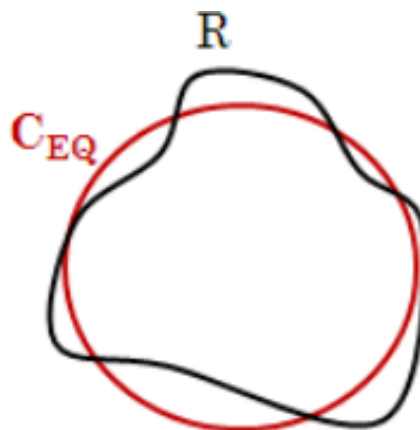
Donc, en utilisant Eq(2.3), Eq(2.4), nous pouvons obtenir le véritable centroïde d'un contour quelque soit le contour normalisé.

### 2.2.1.2 La circularité :

Le Rapport de circularité représente la manière dont une forme est similaire à un cercle [44]. Le rapport de circularité est le rapport entre l'aire d'une forme et l'aire d'un cercle ayant le même périmètre :

$$C_1 = \frac{A_s}{A_c}$$

Équation 2.6



**Figure 2.2 :** Écart de cercle [45].

Les caractéristiques de forme pour la classification des masses mammographiques

Où  $A_s$  est l'aire de la forme et  $A_c$  est l'aire du cercle ayant le même périmètre que la forme. Assumer le périmètre est  $O$ , alors  $A_c = O^2 / 4\pi$ . Ensuite :

$$C_1 = 4\pi \cdot A_s / O^2 \quad \text{Équation 2.7}$$

$4\pi$  est une constante, nous avons donc la deuxième circularité ratio de définition.

- Le rapport de circularité est le rapport entre l'aire d'une forme et son carré périmétrique

$$C_1 = \frac{A_s}{O^2} \quad \text{Équation 2.8}$$

- Le rapport de circularité est également appelé cercle de variance, il est défini comme:

$$C_{va} = \frac{\sigma_R}{\mu_R} \quad \text{Équation 2.9}$$

Où  $\mu_R$  et  $\sigma_R$  sont la moyenne et l'écart type de la distance radiale du centre de la forme  $(g_x, g_y)$  aux points  $P_i$  de contours,  $i \in [0; N - 1]$ . les formules suivantes 2-10 et 2-11 respectivement:

$$\mu_R = \frac{1}{N} \sum_{i=1}^{N-1} d_i \quad \text{et} \quad \sigma_R = \sqrt{\frac{1}{N} \sum_{i=1}^{N-1} (d_i - \mu_R)^2} \quad \text{Équation 2.10}$$

$$O d_i = \sqrt{(x_i - g_x)^2 + (y_i - g_y)^2} \quad \text{Équation 2.11}$$

La forme la plus compacte est un cercle. Voir la figure 2.

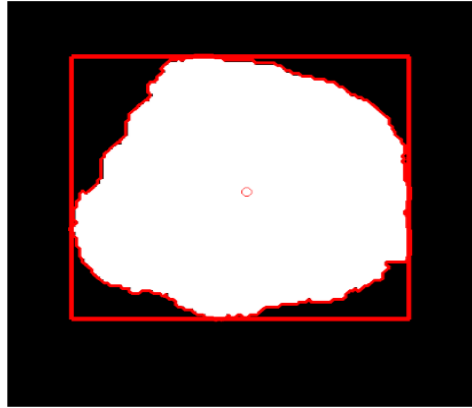
### 2.2.1.3 Rectangularité :

La rectangularité représente le caractère rectangulaire d'une forme, c'est-à-dire la mesure dans laquelle elle correspond à son rectangle de délimitation minimale :

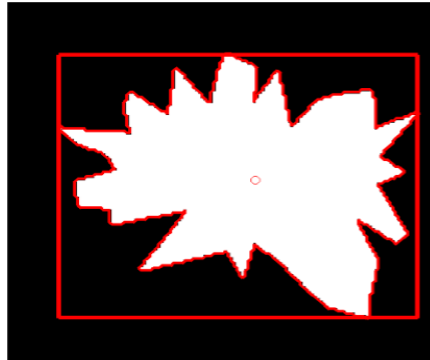
$$\text{Rectangularité} = A_s / A_R \quad \text{Équation 2.12}$$

Les caractéristiques de forme pour la classification des masses mammographiques

Où  $A_s$  est la surface d'une forme;  $A_R$  est la surface du rectangle délimitant le minimum. Selon cette définition, la forme est d'autant plus rectangulaire que sa mesure de rectangularité est proche de 1.



*Figure 2.3: Calcul de rectangularité pour une forme circulaire.*



*Figure 2.4: Calcul de rectangularité pour une forme irrégulière.*

#### 2.2.1.4 La compacité :

C'est une mesure simple de la complexité du contour par rapport à la zone entourée. Elle peut être définie comme suit :

$$Com = \frac{P^2}{A} \quad \text{Équation 1.13}$$

Avec P est le périmètre et A est l'aire de la région entourée. Pour un cercle, la mesure de compacité est minimale et elle est égale à  $4\pi$ . Par contre, une valeur plus élevée sera attribuée pour une forme irrégulière.

Afin de limiter l'intervalle de valeurs prises par cette mesure, Shen et al. [46] ont proposé une que la compacité est calculée par la formule suivante :

$$Com = 1 - \frac{4\pi A}{P^2} \quad \text{Équation 2.14}$$

## Les caractéristiques de forme pour la classification des masses mammographiques

Avec cette expression, la valeur de la compacité tend vers l'unité pour les formes irrégulières et diminue si la forme de la masse est similaire à un cercle.

### 2.2.1.5 L'aire :

C'est le paramètre de forme le plus trivial qui peut être calculé à partir d'un objet détecté sur une image. Il peut être défini comme étant le nombre de pixels contenus dans une lésion. On notera ce paramètre par A [47].

$$A = \sum pixel \in \text{Masse} \quad \text{Équation 2.15}$$

### 2.2.1.6 Le périmètre :

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

$$A = \sum pixel \in \text{Contour} \quad \text{Équation 2.16}$$

### 2.2.1.7 La méthode de coque convexe

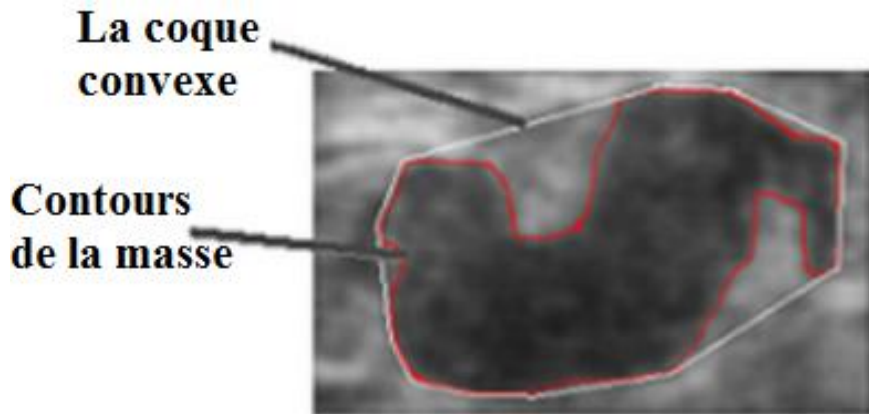
Au cours des dernières années, les chercheurs ont utilisé de nombreux types des caractéristiques pour décrire les formes. Dans l'étude de Behnam et al. [48], six nouvelles caractéristiques de forme ont été proposées :

#### a- Zone de différence

Cette zone est définie par la différence entre la zone de coque convexe et la zone de la masse.

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

La coque convexe est le plus petit polygone convexe qui contient les contours et de la région de la masse, comme il a été représenté sur la Figure 2.5. La zone de la coque convexe est le nombre réel de pixels dans la coque convexe de la masse, et la zone de masse est le nombre de pixels dans la région de masse.



*Figure 2.5 : La coque convexe et le contour d'une masse mammaire [48].*

Cinq caractéristiques peuvent être extraites à travers la fonction *variation* :

#### **b- La Moyenne des variations**

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

$$M_v = \frac{1}{L} \sum_{i=1}^L \text{var}(i) \quad \text{Équation 2.17}$$

#### **c- La variance des variations**

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

$$\text{VAR}_v = \frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^2 \quad \text{Équation 2.18}$$

#### **d- L'asymétrie des variations**

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



Les caractéristiques de forme pour la classification des masses mammographiques

$$SK_v = \frac{\frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^3}{\left(\frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^2\right)^{3/2}} \quad \text{Équation 2.19}$$

#### e- L'aplatissement des variations

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

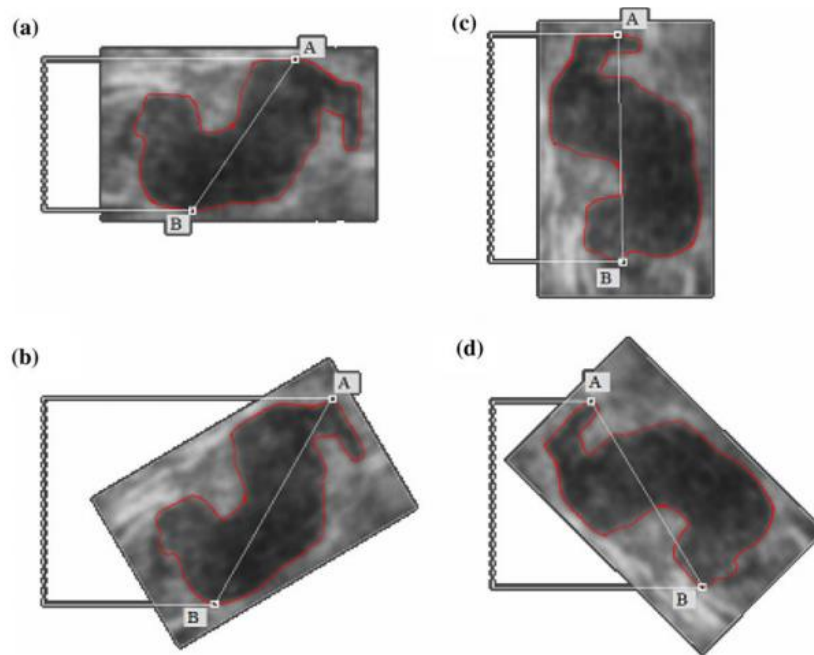
$$APP_v = \frac{(1/L) \sum_{i=1}^L (\text{var}(i) - M_v)^4}{\left((1/L) \sum_{i=1}^L (\text{var}(i) - M_v)^2\right)^2} - 3 \quad \text{Équation 2.20}$$

#### f- L'entropie des variations

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

$$E_v = - \sum_{i=1}^L \text{var}(i) \log_2 \text{var}(i) \quad \text{Équation 2.21}$$

La Figure illustre la base de caractéristiques (**Error! Reference source not found. Équation** ). Pour ces caractéristiques, la sous-image ROI devrait être mise en rotation à différents angles. Rotation avec intervalle de  $5^\circ$  est utilisée. Pour chaque angle des pixels avec des coordonnées minimales Y ('a' dans la figure) et les coordonnées maximales Y ('b' sur la figure) sont trouvés. La projection de la distance entre les pixels sur l'axe Y est calculée et affichée avec une ligne en pointillés dans la figure.



*Figure 2.6: Un exemple de caractéristiques à l'aide d'une masse en rotation à des angles différents [48].*

### 2.3 Processus de classification des masses

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

La structure du système de classification des masses mammographique que nous proposons s'inspire globalement de l'approche du médecin lors de l'examen radiologique pour décrire la forme de la masse mammographique. Notre approche comprend principalement trois étapes : La première phase consiste à la réalisation d'un outil de segmentation manuel basée sur la méthode croissance de région, où les contours de la masse sont délimités par un radiologue expérimenté. La deuxième étape consiste à quantifier les caractéristiques de forme employées par les radiologues selon BI-RADS, dans la dernière étape nous avons utilisé un

## Les caractéristiques de forme pour la classification des masses mammographiques

perceptron multicouches pour classifié les masses mammaires selon leurs formes (Ronde, Ovale, Lobulé ou bien Irrégulière).

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

Dans ce qui suit nous décrivons les méthodes adoptées pour résoudre les étapes principales de notre application, en insistant sur celles qui utilisent les notions présentées dans ce mémoire,

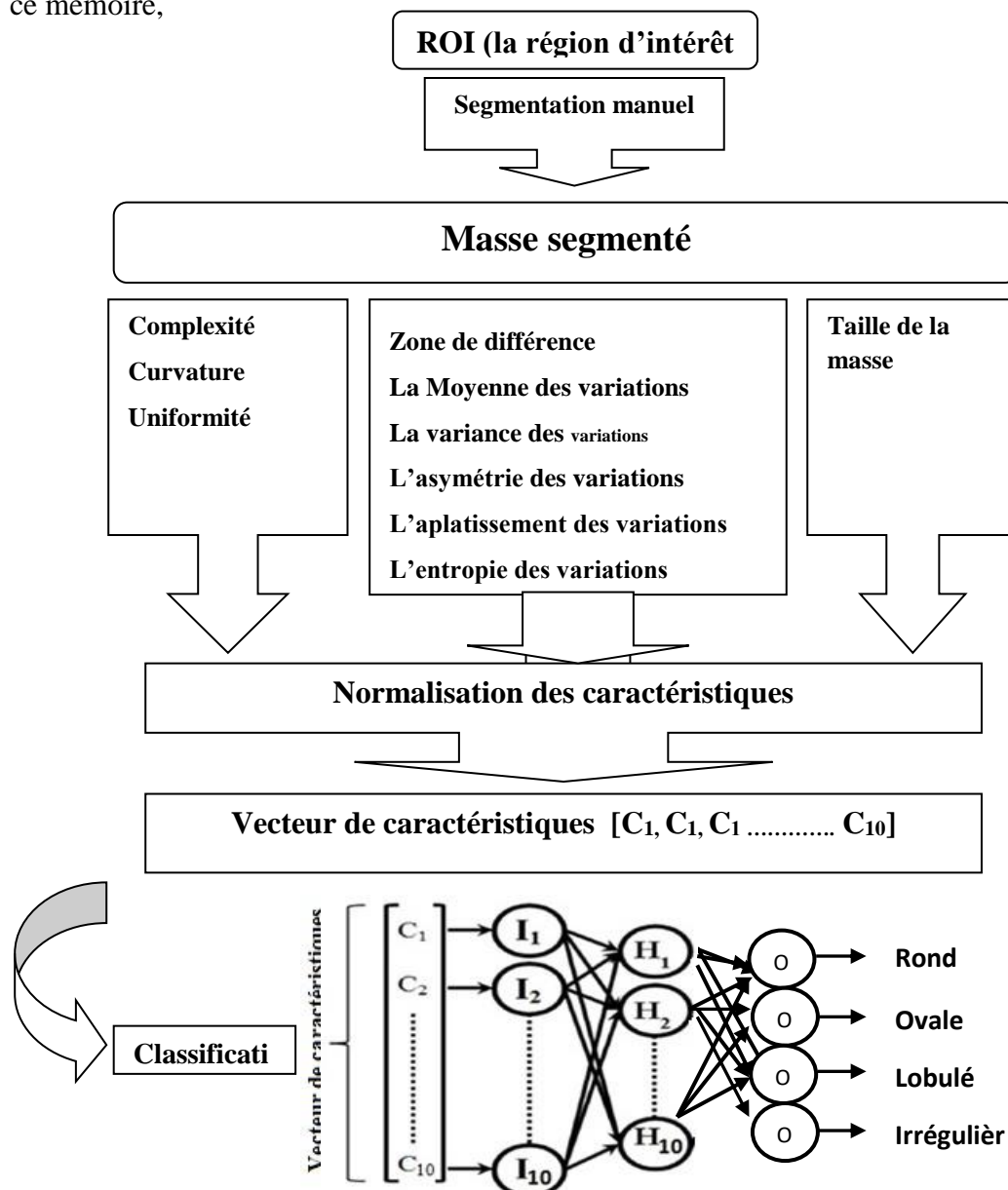


Figure 2.7 : Les étapes de la méthodologie proposée.

### 2.3.1 La sélection de la région d'intérêt (ROI)

Les régions d'intérêt (ROI) sont sélectionnées à partir de la base de données numérique pour la mammographie de dépistage (DDSM) par un radiologue expérimenté pour couvrir toutes les catégories de masses. Les localisations des anomalies dans la mammographie sont disponibles sous forme de chaînes codées dans le fichier « \*. Overlay » (section **Error! Reference source not found.**).

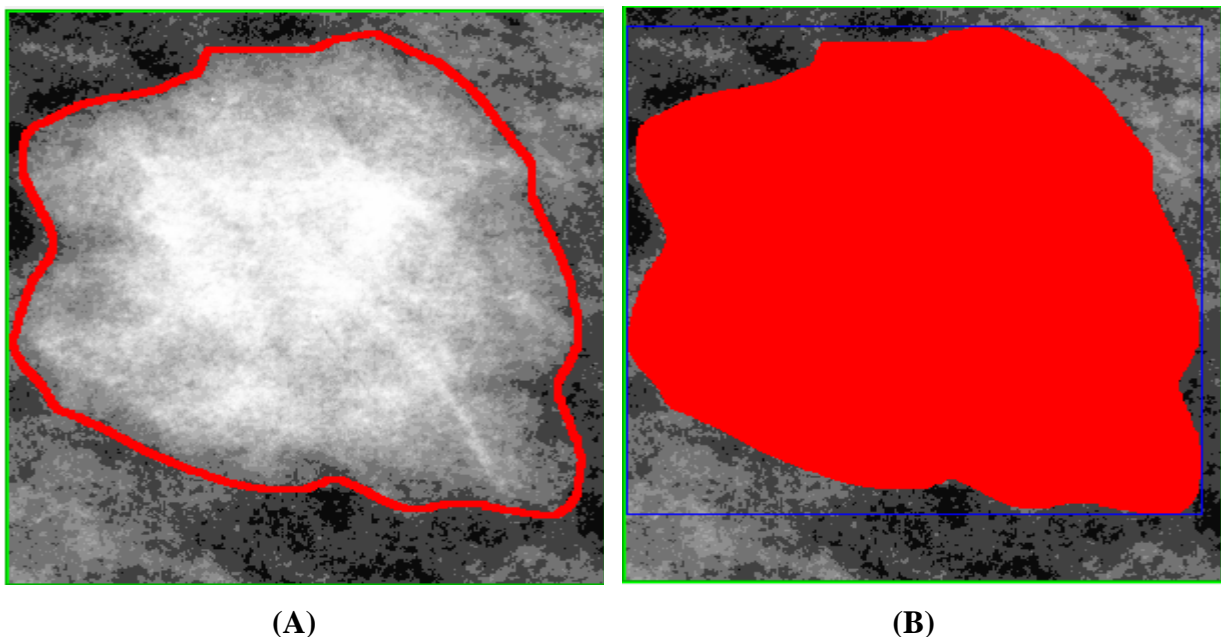
### 2.3.2 La Segmentation des masses mammaires

Dans l'étape de segmentation, les zones suspectées sont séparées du fond. Cette phase est extrêmement importante, car le succès de l'algorithme de classification dépend de cette étape.

Ici les phases d'extraction de caractéristiques et de classification nous intéressent le plus.

Dans notre système, l'étape de segmentation a été faite manuellement par un radiologue expérimenté à travers un outil que nous avons réalisé qui permet de délimiter les bords de la masse mammographique en utilisant la souris **Figure 2.8 (A)**.

Les contours de la masse sont codés sous forme d'une chaîne de chiffres (plus de détail à la section (2.5.2.4) et enregistrés à la base de données. Et la segmentation de la masse a été faite en se basant sur la méthode croissance de région comme le montre la **Figure 2.8 (B)**.



**Figure 2.8 :** La mammographie: A\_1103\_1.LEFT\_CC.LJPEG.1, Masse maligne, Forme: Lobulée.

Les caractéristiques de forme pour la classification des masses mammographiques

La taille de la masse est calculée après cette phase par le nombre de pixels ayant la couleur rouge.

### 2.3.3 L'extraction de caractéristiques des masses

Les descripteurs qui nous intéressent dans ce projet sont les caractéristiques de forme, La

Figure résume les catégories des méthodes de quantification utilisées dans notre approche pour la classification des masses mammaires. Le but de cette section est de quantifier les caractéristiques mentionnées précédemment comme suit :

Avant d'entamer la phase d'extraction des caractéristiques de forme, une étape de mise à l'échelle est nécessaire, cette phase joue un rôle important dans le processus de calcul des caractéristiques de forme; en respectant une échelle de référence permet aux masses identiques d'avoir les mêmes "caractéristiques de forme" quelles que soient leurs dimensions. Lorsque la masse segmentée est mise à l'échelle, l'analyse de la forme peut être effectuée. Cependant, plusieurs caractéristiques de forme ont été proposées dans la littérature. Dans notre travail nous avons utilisé trois caractéristiques : la courbure, la continuité et l'irrégularité extraite de l'énergie interne  $E_{int}$ . Cette dernière a été proposée par Kass et al. Dans [Kass 1988] pour l'utiliser dans leur méthode de détection de contours appelée *Snakes*. Dans leur méthode, l'énergie interne gère la cohérence de la courbe; Il maintient la cohésion des points et la rigidité de la courbe. Mais dans notre travail, nous l'avons utilisée pour décrire la forme de la masse mammographique.

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

Tout d'abord, un ensemble de points  $v_i$  doit être extrait du bord de la masse mammographique pour chaque changement de direction du chemin de contours, la Figure 2.9 montre un exemple du processus de sélection.

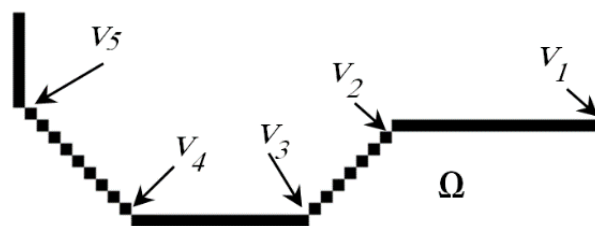


Figure 2.9: Exemple de sélection des points  $v_i$ .

## Les caractéristiques de forme pour la classification des masses mammographiques

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

$$E_{\text{int}}(v) = (\|v'(\Omega)\|^2 + \|v''(\Omega)\|^2) \quad \text{Équation 2.21}$$

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

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

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

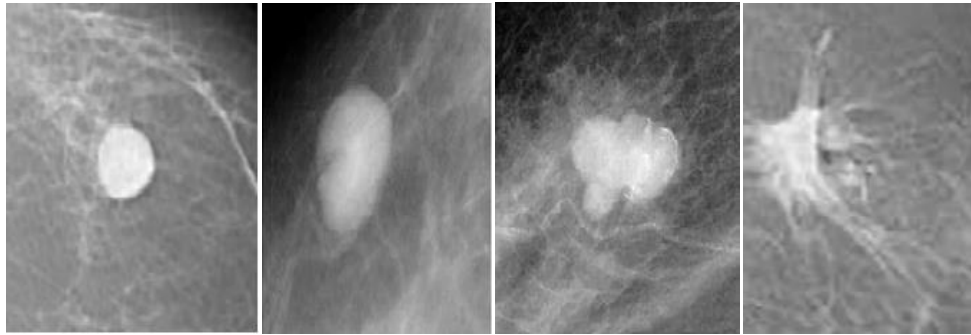
$$\text{Continuité (i)} = \left\| \frac{dv_i}{d\Omega} \right\|^2 \cong |v_i - v_{i-1}|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \quad \text{Équation 2.22}$$

$$\text{Courbure(i)} = \left\| \frac{d^2v_i}{d\Omega^2} \right\|^2 \cong |v_{i-1} - 2v_i + v_{i+1}|^2 = (x_{i-1} - 2x_i + x_{i+1})^2 + (y_{i-1} - 2y_i + y_{i+1})^2 \quad \text{Équation 2.23}$$

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

Les caractéristiques de forme basées sur l'énergie interne	Valeur	
C1= La Continuité	$c_1 = \sum_{i=1}^n \text{Continuité (i)}$	Équation 2.24
C2= La Courbure	$c_2 = \sum_{i=1}^n \text{Courbure (i)}$	Équation 2.25
C3= L'irrégularité	$c_3 = \text{Nombre\_de\_point } s(v_i)$	Équation 2.26

**Tableau 2.1 :** Les Caractéristiques de Forme basées sur l'énergie interne.



a- Ronde

b- Ovale

c- Lobulée

d- Irrégulière

**Figure 2.10:** Les formes de masses (A) Ronde (b) Ovale (c) Lobulée (d) Irrégulière

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

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

Dans ce projet, nous proposons des descripteurs de forme extraites à partir de l'énergie interne du *snake* adaptés au caractère pathologique des masses mammaires. Pour cela, ces descripteurs doivent respecter les critères proposés par Cheikhrouhou et al. [45] :

- ✓ **Invariance** : Une forme peut apparaître de différentes manières et à des échelles différentes, le système visuel humain reconnaîtra presque toujours la même forme. De la même manière, le descripteur de forme doit être invariant par transformations de rotation géométrique.
- ✓ **Unicité** : L'algorithme correspondant au descripteur de forme produit pour chaque forme une description unique ; c'est-à-dire deux formes peuvent produire une même valeur du descripteur, à condition qu'il n'y ait que peu de différences visibles entre ces deux formes.

## Les caractéristiques de forme pour la classification des masses mammographiques

- ✓ **Conservation de l'information** : Il doit être possible de reconstruire la forme d'origine à partir de sa description. Les descripteurs vérifiant ce critère fournissent une représentation complète de la forme.

Nous présentons dans le Tableau les avantages et les inconvénients de chaque descripteur géométrique et la méthode proposée.

<i>Descripteurs</i>	<i>Les avantages</i>	<i>Les inconvénients</i>
<i>A et P</i>	✓ Plus performants quand ils sont associés à d'autres descripteurs	✓ Insuffisants pour décrire les masses complexes ✓ Non uniques
<i>Com</i>	✓ Proportionnel à la complexité de la forme ✓ Invariant	✓ Non unique
<i>C</i>	✓ Simple ✓ Distingue les masses circulaires ✓ Invariant	✓ Insuffisant pour décrire des masses complexes
<i>Rect</i>	✓ Distingue les masses ovales	✓ Non invariant par rotation ✓ Insuffisant pour décrire des masses complexes
<i>Curv</i>	✓ Proportionnel à la complexité de la forme	✓ Non Invariant
Coque convexe	✓ Invariant par rotation ✓ Proportionnel à la complexité de la forme	✓ Insuffisant pour décrire des masses complexes

**Tableau 2.2:** Les avantages et les inconvénients des descripteurs géométriques.

Dans ce projet nous avons utilisé les 06 caractéristiques de la méthode de coque convexe présentées à la section (2.2.1.7) comme le montre le tableau suivant :

<b>Les caractéristiques de forme basées sur la méthode de coque convexe</b>
C4 = La Zone de différence
C5= La Moyenne des variations
C6= La variance des variations
C7=L'asymétrie des variations
C8=L'aplatissement des variations
C9=L'entropie des variations



Les caractéristiques de forme pour la classification des masses mammographiques

*Tableau 2.3 : Les Caractéristiques de Forme basées sur la méthode de coque convexe.*

### 2.3.4 Taille de la masse

En plus de caractéristiques précédemment cités, une caractéristique très utile dans l'étape de classification a été utilisée dans notre système DxAO : la taille de la masse. Ce descripteur est très important pour décrire la forme et peut reclasser une anomalie dans une classe BI-RADS inférieure ou supérieure [49].

Il est facile de calculer la taille réelle de la masse en mm<sup>2</sup> comme le montre (Équation ), qui est représenté par la zone englobé par le contour  $\Omega$ . Où "res" est la résolution utilisée pendant la numérisation.

Autres caractéristiques	Valeur	
C10= Taille de la masse	$(\sum_{p \in \Omega} p) / (res)^2$	Équation 2.27

*Tableau 2.4: Taille de la masse.*

## 2.4 La classification

La notion de classification signifie l'affectation d'une étiquette à des échantillons d'une base de données en utilisant un certain nombre de caractéristiques. Ces caractéristiques doivent bien évidemment être capables d'identifier chaque échantillon

La classification représente la dernière étape dans un système d'aide au diagnostic. Cette étape a pour but d'évaluer la gravité et d'identifier les classes auxquelles appartiennent les anomalies suspectes. Pour ce faire, la méthode adoptée par un système d'aide au diagnostic pour la classification de lésions exploite la description obtenue en sortie de l'étape précédente, à savoir l'extraction de caractéristiques. Les méthodes de classification que nous avons utilisé est la méthode de classification supervisées avec réseaux de neurones artificiel.

### 2.4.1 Réseaux de neurones

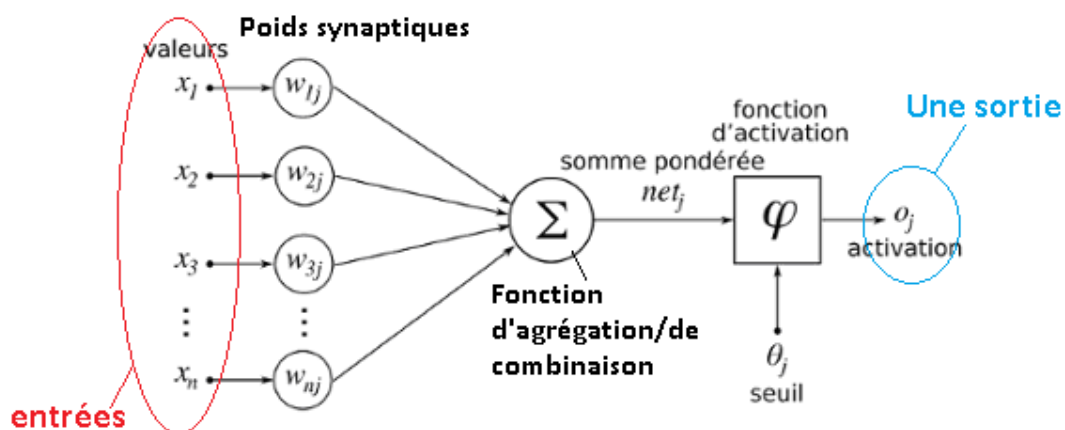
Un réseau de neurones (Artificial Neural Network) est un modèle de calcul dont la conception est schématiquement inspirée du fonctionnement de vrais neurones. Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type statistique, si bien qu'ils sont placés d'une part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

## Les caractéristiques de forme pour la classification des masses mammographiques

Ce sont les deux neurologues *Warren McCulloch* et *Walter Pitts* [51] qui ont mené les premiers travaux sur les réseaux de neurones. Ils constituèrent un modèle simplifié de neurone biologique communément appelé neurone formel. Ils montrèrent également théoriquement que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes.

La fonction des réseaux de neurones formels à l'instar du modèle vrai est de résoudre divers problèmes. A la différence des méthodes traditionnelles de résolution informatique, on ne doit pas construire un programme pas à pas en fonction de la compréhension de celui-ci. Les paramètres les plus importants de ce modèle sont les coefficients synaptiques. Ce sont eux qui construisent le modèle de résolution en fonction des informations données au réseau. Il faut donc trouver un mécanisme, qui permet de les calculer à partir des grandeurs acquises du problème, c'est le principe fondamental de l'apprentissage. Dans un modèle de réseau de neurones formels, apprendre, c'est d'abord calculer les valeurs des coefficients synaptiques en fonction des exemples disponibles. La représentation graphique d'un réseau de neurones artificiel est donnée par la Figure 1.

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



**Figure 2.11:** Représentation d'un neurone artificiel [52].

Les fonctions de seuillage présentent généralement trois intervalles :

1. en dessous du seuil, le neurone est non-actif (souvent dans ce cas, sa sortie vaut 0 ou 1),

## Les caractéristiques de forme pour la classification des masses mammographiques

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

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

Plusieurs types de réseaux de neurones ont été reportés dans la littérature, notamment le perceptron proposé par Rosenblatt [60], les cartes autoorganisatrices de Kohonen [62] et les réseaux basés sur le modèle de Hopfield [61], etc.

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

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

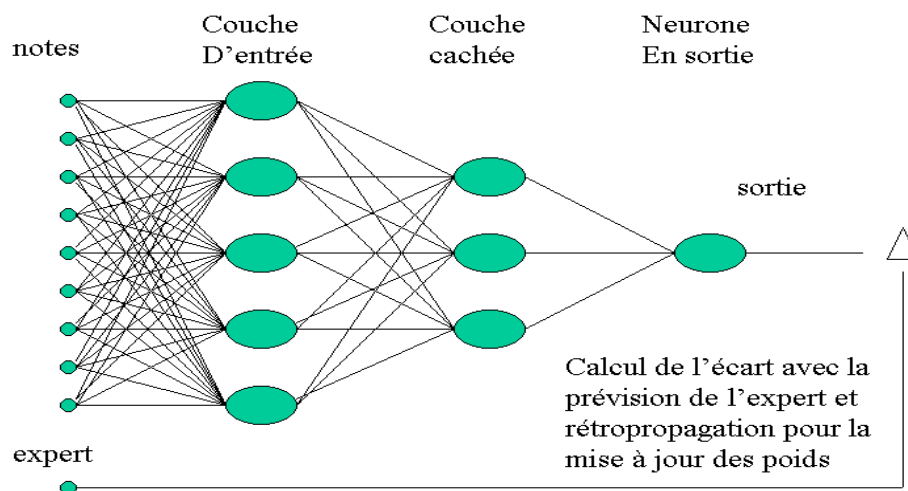
Le PMC est un modèle de réseau à propagation par couche (Figure 2.12). Les neurones y sont organisés en couches successives : une couche d'entrée, une couche de sortie et entre les deux une ou plusieurs couches intermédiaires, appelées aussi couches cachées. Il n'existe pas de connexion entre les neurones d'une même couche, en revanche tout neurone d'une couche est connecté à tous les neurones de la couche suivante. La « couche » d'entrée n'est pas une réelle couche de neurones car elle se contente de coder les variables d'observation. La couche

## Les caractéristiques de forme pour la classification des masses mammographiques

de sortie code la variable de discrimination. Les valeurs d'activité des neurones sont propagées dans le réseau, de l'entrée vers la sortie, sans retour arrière. La présence d'une couche cachée permet de modéliser des relations non linéaires entre les entrées et la sortie. En théorie une seule couche cachée suffit, mais le fait de disposer d'une seconde couche cachée permet de modéliser plus facilement une fonction de discrimination non continue. En pratique, la plupart des problèmes sont résolus avec un ou deux niveaux, trois au maximum.

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

Dans le cas du PMC, on utilise un apprentissage supervisé. Les valeurs des poids de connexion sont créées tout d'abord au hasard et le système cherche par itérations successives à obtenir une modélisation des données. A chaque étape, une entrée est présentée au réseau, il propage ces valeurs vers les neurones de sortie. Cette sortie calculée est comparée avec la réponse attendue et le système modifie les poids en conséquence. Cette altération des connexions est obtenue par l'algorithme de rétropropagation du gradient d'erreur. Ce calcul est chargé de rétropropager dans le réseau les erreurs constatées sur les sorties. En théorie, on ne peut jamais être sûr que cet algorithme finisse par déterminer un ensemble de poids convenable pour tous les couples d'entrées-sorties. En pratique, on ne construit pas un seul RNA, mais plusieurs modèles en jouant sur les paramètres de cet algorithme, et en cherchant à obtenir un modèle qui colle au mieux aux données.



**Figure 2.12:** Schéma d'un perceptron multicouches (PMC) [51].

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

Les réseaux de neurones sont célèbres pour leurs bonnes performances, en termes de taux et de vitesse de classification. Ils ont aussi l'avantage de fonctionner en présence de données manquantes ou fausses. Ces avantages font la popularité des approches neuronales en reconnaissance de formes [53].

L'inconvénient majeur des réseaux de neurones est l'absence de méthode systématique permettant de définir la meilleure architecture du réseau et le nombre de neurones à placer pour chaque couche cachée, il est difficile de choisir l'architecture et de régler les paramètres d'apprentissage. En effet, le résultat de l'apprentissage est un réseau défini par une fonction d'activation et un très grand nombre de poids à valeurs réelles. Ces poids sont difficilement interprétables. De plus les réseaux de neurones ont un côté « boîte noire » : les résultats de la classification sont difficilement interprétables.

#### 2.4.2 La classification utilisée

La classification dans notre approche a été faite en utilisant un PMC (Perceptron Multicouches). PMC est un réseau neuronal artificiel (RNA) avec apprentissage supervisé, à savoir, l'algorithme de rétro-propagation, spécialement dédié aux fins de classification. Il a la capacité de séparer des classes de motifs non linéairement séparables. Notre vecteur de caractéristiques est composé de 10 caractéristiques extraites à l'étape précédente,

## Les caractéristiques de forme pour la classification des masses mammographiques

En effet, le processus de classification consiste à classer les masses sur les quatre classes (Ronde, Ovale, Lobulé et Irrégulière) ce qui signifie que le nombre neurones de la couche de sortie  $O = 4$ .

Dans ce travail le nombre de neurones d'entrée est le même que la longueur du vecteur de caractéristiques calculé dans l'étape précédente ( $I = 10$ ).

Comme mentionné dans [54] une couche cachée est suffisante pour résoudre tout problème complexe non linéaire. Pour déterminer le nombre de neurones dans la couche cachée, nous nous basons sur une heuristique proposée par Berry et al. [55] Qui estiment que le nombre de neurones de couche cachée ' $H$ ' doit être inférieur à deux fois le nombre de neurones dans la couche d'entrée. Le nombre de neurones cachés  $H$  est calculé pour notre cas en utilisant la règle :  $H=O+ (0.75 \times I)$  et  $H < 2 \times I$ . Cela signifie que : Dans notre Scénario  $H= 12$ ,

Toutes les caractéristiques d'entrée prennent des valeurs qui peuvent être très éloignées les unes des autres. Par exemple, la caractéristique courbure peut avoir une valeur maximale de 1500, et la continuité peut avoir un maximum de 1000. Pour éviter cela, nous devons normaliser les valeurs d'entrée en utilisant la formule suivante:

$$f_{Nor} = (f - \text{Min}(f)) / (\text{Max}(f) - \text{Min}(f)) \quad \text{Équation 2.28}$$

Où:

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

$f_{Nor}$  : La caractéristique normalisée.

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

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

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

## 2.5 La base de données mammographiques

### 2.5.1 Mammographic Image Analysis Society (MIAS)

C'est une base de données de mammographies numériques : Elle contient 322 images MLO rangées en paires [56]: celles dont le numéro est pair correspondent aux MLO gauches et

Les caractéristiques de forme pour la classification des masses mammographiques

celles dont le numéro est impair sont les MLO droites. Toutes les images ont une taille de 1024x1024 pixels.

### **2.5.2 Digital Database for Screening Mammography (DDSM)**

La base de données numérique pour la mammographie de dépistage (DDSM) [58] est une base de données de mammographies qui a été rassemblée par un groupe de chercheurs de l'Université Sud De Florida en collaboration avec l'hôpital "Massachusetts General Hospital" (MGH), l'hôpital Washington University of St. Louis School of Medicine" (WUSTL) et l'université "Wake Forest University" (WFU).

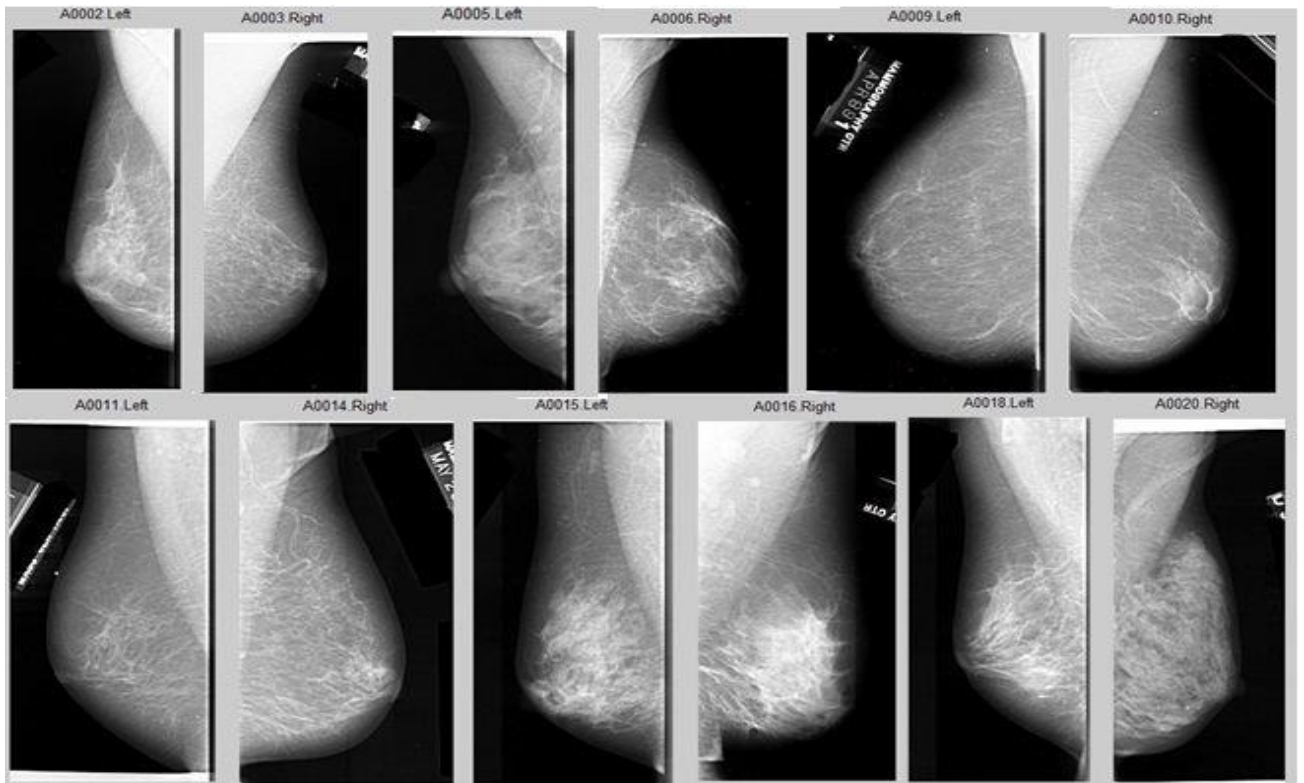
Cette base de données contient 2620 cas correspondant à chacune une patiente différente. Chaque cas contient les vues obliques externe (MLO) et face externe (CC) de chaque sein.

Et de métadonnées associées à chaque cas, elles sont contenues dans le fichier ics et les fichiers Overlay

Chaque dossier de patiente est composé de :

- \_ 1 fichier .ics décrivant en format ASCII, les informations générales d'un dossier de patient.
- \_ 4 fichiers images .LJPEG (LOSSLESS JPEG) des radios numérisées.
- \_ Chaque radio présente un angle de vue du sein : Left-CC, Left-MLO, Right-CC, Right-MLO (CC : Cranio Caudal ,MLO : Medio Latral Oblique).
- \_ Pour chaque radio présentant une ou des zones anormales, est associé un fichier OVERLAY en format ASCII, décrivant une anomalie du sein.
- \_ Fichier image .16-PGM regroupant les 4 radios et présentant un aperçu rapide pour la visualisation d'un dossier de patient.

## Les caractéristiques de forme pour la classification des masses mammographiques



*Figure 2.13 : Des échantillons de la base DDSM [57]*

Les images contenant des zones suspectes sont associées à des informations sur les emplacements et les types de régions suspectes. Trois cas de diagnostic sont possibles : les cas normaux (absence de lésions), les cas bénins (existence de lésions bénignes) et les cas malins (existence de lésions malignes), Cette description contient le nombre et le type d'anomalies (masse ou micro-calcification), l'emplacement de ces anomalies ainsi que l'évaluation selon le code BI-RADS. Le tableau 2.5 montre la répartition de la base DDSM selon le diagnostic.

<b>Diagnostic</b>	<b>Nombre d'images</b>
Normaux	695 images
Bénins	870 images
Bénins sans rappel	141 images
Cancer	914 images

*Table 2.5 : Répartition des images de la base DDSM selon le diagnostic*



### 2.5.2.1 Le fichier .ics (Exemple : B-3024-1.ics)

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

Figure 2.14 : Les informations dans un fichier ics

### 2.5.2.2 Description du fichier .ics

Le fichier .ics contient des informations importantes En format ASCII telles que la date de l'étude, l'âge du patient, la date de la numérisation des films, le type de numériseur utilisé, une densité du tissu mammaire, la taille de chaque fichier image, le nombre de bits par pixel, la résolution de la numérisation. Le fichier ".ics" donne également une cote de densité de tissus mammaires de 1 à 4 selon l'évaluation ACR réalisé par un radiologue expérimenté. les quatre images "LJPEG" ont des "Overlay" et ces fichiers sont répertoriés dans la **Error! Reference source not found.** Si les lignes de description d'image avaient "NON-OVERLAY" au lieu de "OVERLAY", les mammographies n'auraient pas de fichiers « \*. Overlay ». Comme illustre la figure 2.15.

### 2.5.2.3 Description des fichiers ".LJPEG" :

Les images ont toutes été stockées dans un format utilisant la compression LOSSLESS JPEG. Même avec la compression, chaque fichier image est très volumineux car les films ont été scannés avec une résolution comprise entre 42 et 100 microns. Une version exécutable pour SunOS 5.5 est utilisée pour décompresser les images. Une fois décompressé, chaque fichier image ne contient que des valeurs brutes de pixels. Parce qu'il n'y a pas d'information d'en-tête dans le fichier, la taille de chaque image doit être obtenue à partir du fichier ".ics".

### 2.5.2.4 Description du fichier overlay

## Les caractéristiques de forme pour la classification des masses mammographiques

Chaque anomalie contient des informations sur le type de lésion, l'évaluation, la subtilité, la pathologie et au moins un contour. Les mots-clés qui décrivent le type de lésion sont tirés du lexique ACR BI-RADS. Le code d'évaluation est une valeur de 1 à 5. (Figure 2.15)

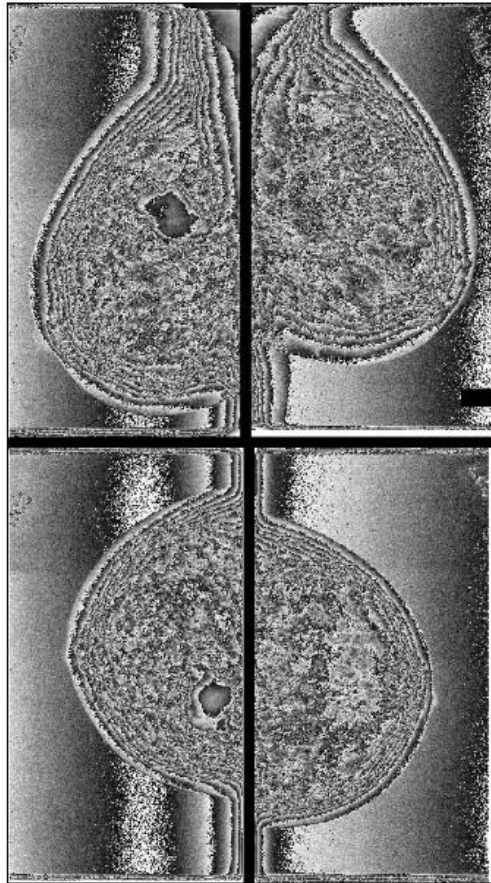
### Le fichier B\_3024\_1.RIGHT\_MLO.OVERLAY

```
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE MASS SHAPE OVAL MARGINS ILL_DEFINED
ASSESSMENT 4
SUBTLETY 3
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 1
BOUNDARY 1284 4076 0 0 0 0 0 0 0 0 . . . 6 6 6 6 0 0 #
```

*Figure 2.15 : Un exemple d'un fichier overlay*

#### 2.5.2.5 La description des fichiers « .16\_PGM »

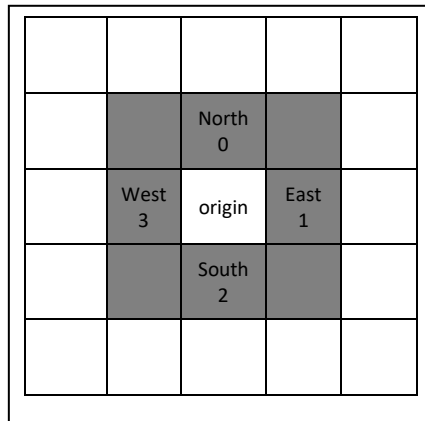
Les fichiers ".16\_PGM" sont des images sous-échantillonnées concaténées. Ils sont stockés dans un format PGM (Portable Gray Map) sur 16 bits. Le programme populaire "imageJ" pouvait être utilisé pour visualiser ces images. Un exemple du résultat de l'affichage est présenté sur la Figure 6. La qualité de l'image affichée peut être médiocre même si le fichier contient des informations assez bonnes. Le but de ce type de fichiers est simplement de fournir une vue "rapide" sur les quatre images mammographiques LJPEG.



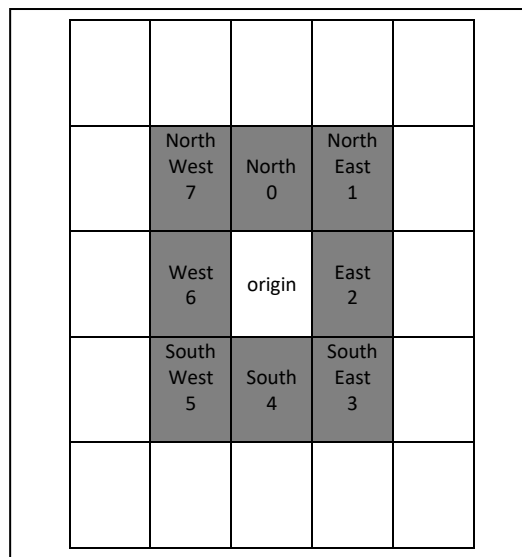
**Figure 2.16:** *La visualisation d'un fichier ".16\_PGM".*

#### 2.5.2.6 Codage en chaîne

Afin d'obtenir une représentation d'un contour, on peut simplement enregistrer les coordonnées de la séquence de pixels dans l'image. Alternativement, nous pouvons aussi enregistrer la position relative entre des pixels consécutifs. C'est l'idée de base derrière les codes de chaîne. Une chaîne codée est l'une des plus anciennes techniques de vision par ordinateur introduites à l'origine dans les années 1960 par Freeman. Essentiellement, l'ensemble des pixels de la bordure d'une forme est traduit en un ensemble de connexions entre eux. Étant donné une bordure complète, celle qui est un ensemble de points connectés, alors à partir d'un pixel, nous devons être en mesure de déterminer la direction dans laquelle le pixel suivant doit être trouvé, à savoir, le pixel suivant est l'un des points adjacents dans une des directions principales de la boussole. Ainsi, le code de chaîne est formé en concaténant le nombre qui désigne la direction du pixel suivant devient un élément dans le code final. Ceci est répété pour chaque point jusqu'à ce que le point de départ soit atteint lorsque la forme (fermée) est complètement analysée.



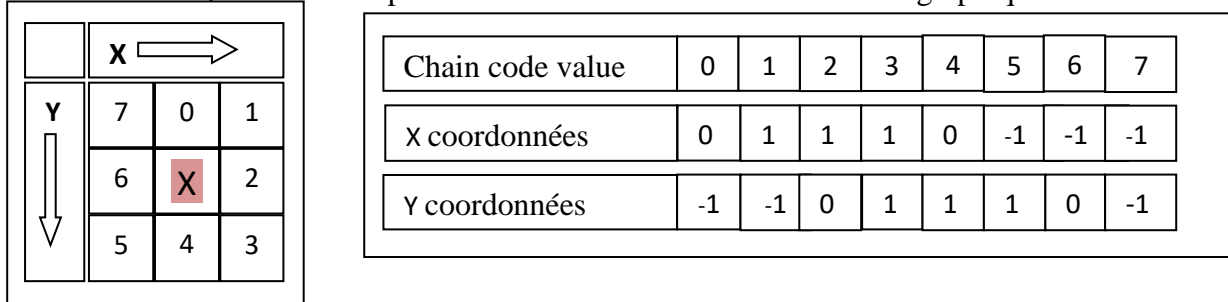
*Figure 2.17 :Connectivité 4 voies*



*Figure 2.18 : Connectivité 8 voies*

Pour toutes les mammographies contenant une lésion, le spécialiste radiologue a tracé un contour autour de la région tumorale, ce contour est analysé sous forme de chaîne à l'aide du code de Freeman. Les deux premières valeurs de la chaîne représentent la colonne et la ligne de départ du code de chaîne de cet ordre, sur lesquels on se base pour déterminer les coordonnées des autres pixels du contour et un caractère "#" indique la fin du code de chaîne. Les chiffres correspondent aux indications suivantes :

## Les caractéristiques de forme pour la classification des masses mammographiques



**Figure 2.19 :** la représentation des valeurs de déplacement de la chaîne codée.

Nous avons utilisé cette technique aussi pour la codification des contours tracés par le radiologue, ce qui nous facilitera par la suite l'enregistrement des contours à la base de données.

## 2.6 Environnement et outils de développement

### 2.6.1 Plateformes utilisées

Notre application a été implémentée sous Windows 8.1, Les clichés mammographique sont décompressés sous linux elementary OS

### 2.6.2 Langage de programmation et outils

#### 2.6.2.1 Java

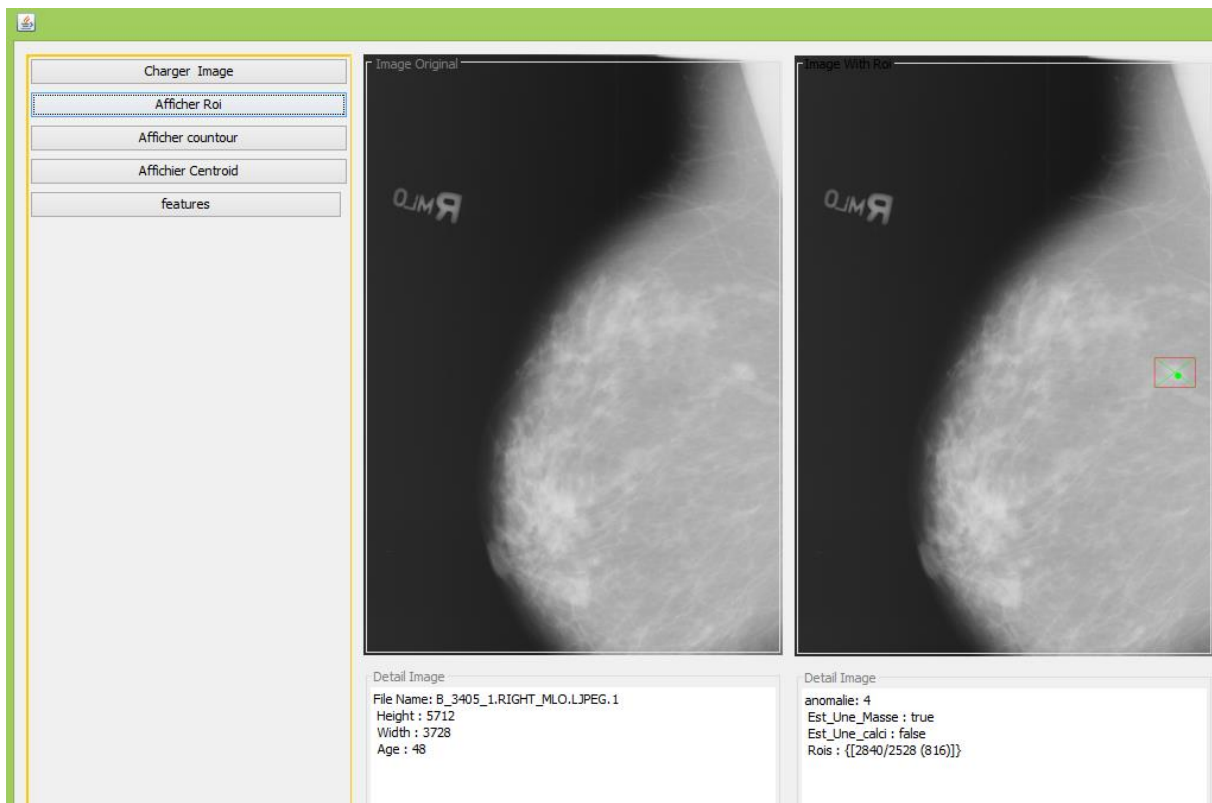
Le langage de programmation utilisé dans notre application est JAVA, en utilisant l'API java AWT et la couche accès aux données JPA (Java Persistence Api) en utilisant la plateforme eclipselink pour stocker les objets java dans un environnement de production de logiciels eclipse oxygène qui s'appuyant principalement sur Java.

Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Beaucoup d'applications et de sites Web ne fonctionnent pas si Java n'est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable. Des ordinateurs portables aux centres de données, des consoles de jeux aux superordinateurs scientifiques, des téléphones portables à Internet, la technologie Java est présente sur tous les fronts.

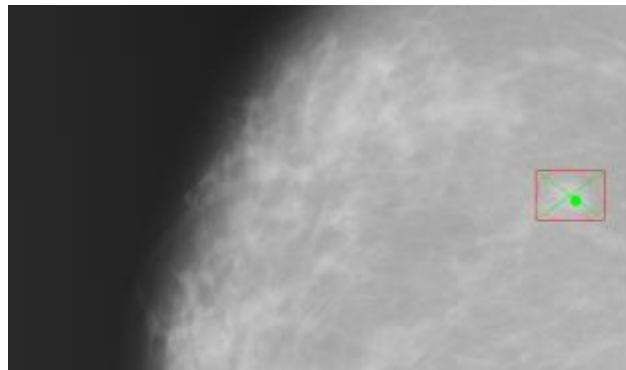
#### 2.6.2.2 Eclipse

Eclipse est un IDE, Integrated Development Environment (EDI environnement de développement intégré en français), c'est-à-dire un logiciel qui simplifie la programmation en proposant un certain nombre de raccourcis et d'aide à la programmation. Il est développé par IBM, est gratuit et disponible pour la plupart des systèmes d'exploitation.

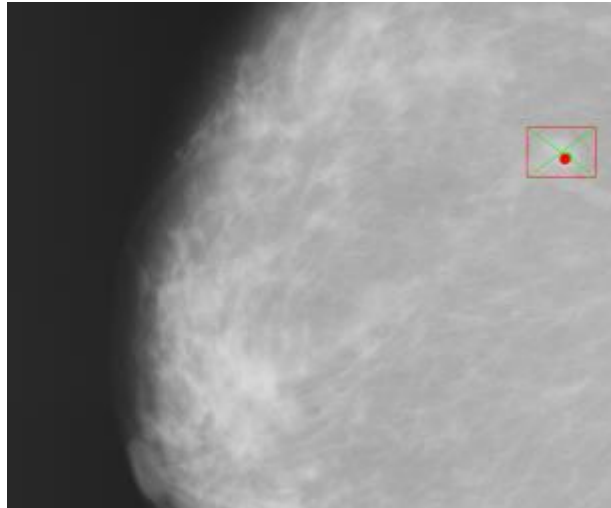
## 2.7 Réalisation et interface



*Figure 2.20 : affichage de l'image original avec la région d'intérêt*



*Figure 2.21 : affichage de contour de la pathologie*



*Figure 2.22 : affichage de centroid de la pathologie*

## 2.8 Résultats et discussions

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

Le système DAOx résumé Dans le cadre d'une classification supervisée, la sous base sélectionnée à partir de la base DDSM est partagée en 2 parties. Une partie composée de 282 images est consacrée à l'apprentissage et à la création du modèle et une partie composée de 187 images est consacrée à la phase de test de la performance du système de diagnostic (voir tableau 2.6). Le choix des bases d'apprentissage et de test est aléatoire. Afin d'obtenir un modèle robuste

La Forme	Ronde	Ovale	Lobulée	Irrégulière	Total
	88	123	93	165	469
Apprentissage	53	74	56	99	282
validation	35	49	37	66	187

**Tableau 2.6 :** Les masses utilisées.

Les expérimentations ont été faites sur un ensemble de 469 masses choisies à partir de la base de données mammographique (DDSM) [25]. Dans DDSM, les mammographies ont été scannées en utilisant une résolution entre 0,042 et 0,05 mm comme taille de pixel, avec une taille d'environ  $3000 \times 5000$  pixels. Chaque anomalie contient des informations sur le type de lésion et sa forme, l'évaluation BI-RADS, la pathologie et les régions d'intérêt (ROI) de la masse suspectée. Les anomalies utilisées dans cette étude sont choisies par un radiologue

Les caractéristiques de forme pour la classification des masses mammographiques

expérimenté pour toucher toutes les catégories de forme; et répartis également sur les quatre classes de formes (Ronde, ovale, lobulaire et irrégulière), comme le montre Tableau 2.6.

La matrice de confusion du résultat est présentée dans le tableau suivant :

Classes		Classifié Comme Forme				Taux de classification
		Ronde	Ovale	Lobulée	Irrégulière	
La Forme	Ronde	24	7	3	1	68.57%
	Ovale	5	38	4	2	77.55%
	Lobulée	2	6	25	4	67.55%
	Irrégulière	3	5	8	50	75.75%
Bien classifié		137 / 187				<b>73.26 %</b>

**Tableau 2.7:** La matrice de confusion pour la classification des anomalies selon leurs forme.

## 2.9 Conclusion

Au cours de ce chapitre, nous avons présenté l'architecture de notre système et les résultats obtenus. Les images utilisées ont été segmenté par un radiologue expérimenté pour toucher toutes les catégories de formes ; dans cette architecture, nous avons utilisé 10 caractéristiques de forme pour classer une masse mammographique dans l'une des quatre Forme (Ronde, Ovale, Lobulée et Irrégulière). Un taux de reconnaissance global égale à 73,26 en utilisant 469 images mammographiques.



# Conclusion générale

Dans ce travail de mémoire, on s'est intéressé à l'étude des systèmes automatiques de diagnostic assistés par ordinateur en vue du diagnostic du cancer du sein, notamment la description et la classification des masses mammaires selon leurs formes. En effet, les statistiques confirment que le cancer du sein représente une menace prépondérante pour la vie de la femme. Cependant, une telle menace n'est gérée que par la prise en charge rapide de la maladie pour maximiser les chances de survie. D'où, le recours au moyen informatique automatisant la procédure de diagnostic appelé le Diagnostic Assisté par Ordinateur (DxAO). Cette procédure permet d'assister les radiologues dans l'analyse des mammographies dont le nombre ne cesse de croître ces dernières années. Mener à bien ce travail de recherche à nécessité tout d'abord la maîtrise de certaines connaissances dans le domaine du cancer du sein. Les recherches dans le domaine de l'image se focalisent sur différentes étapes de la chaîne du diagnostic assisté par ordinateur afin de donner une description détaillée de l'anomalie.

Un état concernant les étapes d'une chaîne de diagnostic assisté par ordinateur a été réalisé afin de réussir la description et la classification des masses mammaires. Dans ce contexte, un outil de segmentation manuel a été réalisé et utilisé par un radiologue pour délimiter les contours des anomalies mammaires.

Dans le but d'analyser la forme des masses, Une étape de description des formes obtenues à partir de l'étape de segmentation pour caractériser les masses, 10 caractéristiques ont été calculées en se basant sur l'énergie interne du contour actif et la méthode de coques convexes. La dernière étape c'est l'étape de classification qui permet de classifier les masses selon leurs formes (Ronde, Ovale, Lobulaire ou bien Irrégulière).

Il est nécessaire également de chercher d'autres caractéristiques plus efficaces et aussi d'utiliser d'autres classifieurs afin d'améliorer le taux de reconnaissance.

# Bibliographie

- [1] Majid, A. S., de Paredes, E. S., Doherty, R. D., Sharma, N. R., & Salvador, X. (2003). Missed breast carcinoma: pitfalls and pearls. *Radiographics*, 23(4), 881-895.
- [2] <https://www.passeportsante.net/fr/parties-corps/Fiche.aspx?doc=sein> [Accès le 2019 0422].
- [3] [www.contrelecancer.ma/site\\_media/uploaded\\_\\_les/Guide\\_de\\_detection\\_pre%3%Bccoce\\_des\\_cancers\\_du\\_sein\\_et\\_du\\_col\\_de\\_lute%3%BCrus.pdf](http://www.contrelecancer.ma/site_media/uploaded__les/Guide_de_detection_pre%3%Bccoce_des_cancers_du_sein_et_du_col_de_lute%3%BCrus.pdf) [Accès le 2019 04 22].
- [4] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Les-points-cles> [Accès le 2019 04 20].
- [5] <http://www.afblum.be/bioafb/cancsein/cancsein.htm>. [Accès le 2019 05 07].
- [6] HAMDI, N., & MERZOUK, K. (2017). *Etude épidémiologique et Anatomo-pathologique Du cancer du sein dans la région de Bouira* (Doctoral dissertation, Université de Bouira).
- [7] Stinès, J. (2012). *Imagerie du sein*. Elsevier Health Sciences.
- [8] <https://www.everzen.fr/survie-cancer-02021069.htm>"EverZen: Quel est le taux de survie du cancer du sein?" [Accès le 2019 04 20].
- [9] Tamimi, R. M., Byrne, C., Colditz, G. A., & Hankinson, S. E. (2007). Endogenous hormone levels, mammographic density, and subsequent risk of breast cancer in postmenopausal women. *Journal of the National Cancer Institute*, 99(15), 1178-1187.
- [10] Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). Screening for breast cancer. *Jama*, 293(10), 1245-1256.
- [11] Le-Petross, C. H., Bidaut, L., & Yang, W. T. (2008, February). Evolving role of imaging modalities in inflammatory breast cancer. In *Seminars in oncology* (Vol. 35, No. 1, pp. 51-63). WB Saunders.
- [12] Sardanelli, F., Podo, F., D'Agnolo, G., Verdecchia, A., Santaquilani, M., Musumeci, R., ... & Federico, M. (2007). Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. *Radiology*, 242(3), 698-715.

- [13] Namer, M., Héry, M., Serin, D., Spielmann, M., & Gligorov, J. (2008). *Cancer du sein: Compte rendu du cours supérieur francophone de cancérologie, (Saint-Paul-de-Vence, 18-20 janvier 2007)*. Springer Science & Business Media.
- [14] Bovis, K., & Singh, S. (2000, September). Detection of masses in mammograms using texture features. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*(Vol. 2, pp. 267-270). IEEE.
- [15] Beam, C. A., Sullivan, D. C., & Layde, P. M. (1996). Effect of human variability on independent double reading in screening mammography. *Academic radiology*, 3(11), 891-897.
- [16] Vyborny, C. J., Giger, M. L., & Nishikawa, R. M. (2000). Computer-aided detection and diagnosis of breast cancer. *Radiologic Clinics of North America*, 38(4), 725-740.
- [17] Hall, F. M. (1997). Opinions about mammographic double reading. *AJR. American journal of roentgenology*, 168(3), 846-846.
- [18] Hall, F. M. (1997). Opinions about mammographic double reading. *AJR. American journal of roentgenology*, 168(3), 846-846.
- [19] Liberman, L., Morris, E. A., Dershaw, D. D., Abramson, A. F., & Tan, L. K. (2003). MR imaging of the ipsilateral breast in women with percutaneously proven breast cancer. *American Journal of Roentgenology*, 180(4), 901-910.
- [20] Sampat, M. P., Markey, M. K., & Bovik, A. C. (2005). Computer-aided detection and diagnosis in mammography. *Handbook of image and video processing*, 2(1), 1195-1217.
- [21] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P., & Du, H. N. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern recognition*, 39(4), 646-668.
- [22] BI-RADS, A. C. R. (2003). breast imaging and reporting data system. breast imaging atlas; mammography, breast ultrasound, magnetic resonance imaging Reston. Va: American College of Radiology.
- [23] Boyd, N. F., Jensen, H. M., Cooke, G., & Han, H. L. (1992). Relationship between mammographic and histological risk factors for breast cancer. *JNCI: Journal of the National Cancer Institute*, 84(15), 1170-1179.
- [24] Boyd NF, Dite GS, Stone J, Gunasekara A, English DR, et al. (2002) Heritability of mammographic density, a risk factor for breast cancer. *N Engl J Med* 347: 886-894.

- [25] D'Orsi, C. J. (1996). The American College of Radiology mammography lexicon: an initial attempt to standardize terminology. *AJR. American journal of roentgenology*, 166(4), 779-780.
- [26] CHEKKAF, H., & TOUIL, I. *Segmentation des masses dans les images mammographiques* (Doctoral dissertation).
- [27] McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiology and Prevention Biomarkers*, 15(6), 1159-1169.
- [28] Ursin, G., Ma, H., Wu, A. H., Bernstein, L., Salane, M., Parisky, Y. R., ... & Pike, M. C. (2003). Mammographic density and breast cancer in three ethnic groups. *Cancer Epidemiology and Prevention Biomarkers*, 12(4), 332-338.
- [29] Cheikhrouhou, I. (2012). *Description et classification des masses mammaires pour le diagnostic du cancer du sein* (Doctoral dissertation).
- [30] Jing, H., Yang, Y., & Nishikawa, R. M. (2010). Detection of clustered microcalcifications using spatial point process modeling. *Physics in Medicine & Biology*, 56(1), 1.
- [31] Bird, R. E., Wallace, T. W., & Yankaskas, B. C. (1992). Analysis of cancers missed at screening mammography. *Radiology*, 184(3), 613-617.
- [32] Burrell, H. C., Sibbering, D. M., Wilson, A. R., Pinder, S. E., Evans, A. J., Yeoman, L. J., ... & Robertson, J. F. (1996). Screening interval breast cancers: mammographic features and prognosis factors. *Radiology*, 199(3), 811-817.
- [33] Eziddin, W. (2012). *Segmentation itérative d'images par propagation de connaissances dans le domaine possibiliste* (Doctoral dissertation, Thèse, Télécom Bretagne).
- [34] Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E. R., & Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical image analysis*, 14(2), 87-110.
- [35] P. J.F.HERON, «oncoprof.net,» [En ligne]: [http://www.oncoprof.net/Generale2000-/g04\\_Diagnostic/Mammographie/g04\\_mm05.html](http://www.oncoprof.net/Generale2000-/g04_Diagnostic/Mammographie/g04_mm05.html). [Accès le 18 05 2019].
- [36] Wolfe, J. N. (1976). Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, 126(6), 1130-1137.;
- [37] Obenauer, S., Fischer, U., Baum, F., Dammert, S., Füzesi, L., & Grabbe, E. (2001). Stereotactic vacuum core biopsy of clustered microcalcifications classified as B1-RADS

3. *RoFo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*, 173(8), 696-701.

[38] Bérubé, M., Curpen, B., Ugolini, P., Lalonde, L., & Ouimet-Oliva, D. (1998). Level of suspicion of a mammographic lesion: use of features defined by BI-RADS lexicon and correlation with large-core breast biopsy. *Canadian Association of Radiologists journal= Journal l'Association canadienne des radiologistes*, 49(4), 223-228.

[39] Orel, S. G., Kay, N., Reynolds, C., & Sullivan, D. C. (1999). BI-RADS categorization as a predictor of malignancy. *Radiology*, 211(3), 845-850.

[40] Lo, J. Y., & Floyd, C. E. (2000, November). Predicting malignancy and invasion of breast lesions from BI-RADS and history data: New results from 1027 cases.

In *Radiology* (Vol. 217, pp. 317-317). 20TH AND NORTHAMPTON STS, EASTON, PA 18042 USA: RADIOLOGICAL SOC NORTH AMER..

[41] Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern recognition*, 37(1), 1-19.

[42] Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis: Wiley series in probability and statistics*.

[43] Costa, L. D. F. D., & Cesar Jr, R. M. (2000). *Shape analysis and classification: theory and practice*. CRC Press, Inc..

[44] Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern recognition*, 37(1), 1-19.

[45] Cheikhrouhou, I., Djemal, K., Sellami-Masmoudi, D., Derbel, N., & Maaref, H. (Octobre 2007). *Abnormalities description for breast cancer recognition*. International Conference on E-medical Systems (pp. 199-205). Fez (Maroc).

[46] Shen, L., Rangayyan, R. M., & Desautels, J. L. (1993). Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(06), 1403-1416.

[47] Bottigli, U., Cascio, D., Fauci, F., Golosio, B., Magro, R., Masala, G. L., ... & Stumbo, S. (2006, March). Massive lesions classification using features based on morphological lesion differences. In *Proceedings of World Academy of Science, Engineering and Technology* (Vol. 12, pp. 20-24).

- [48] Behnam, H., Zakeri, F. S., & Ahmadinejad, N. (2010). *Breast mass classification on sonographic images on the basis of shape analysis*. *Journal of Medical Ultrasonics*, 37 (4), 181-186.
- [49] Lévy, L., Suissa, M., Bokobsa, J., Tristant, H., Chiche, J. F., Martin, B., & Teman, G. (2005). *Présentation de la traduction française du BI-RADS®(Breast Imaging Reporting System and Data System)*. *Gynécologie obstétrique & fertilité*, 33 (5), 338-347.
- [50] Kass, G. V. (1980). *An exploratory technique for investigating large quantities of categorical data*. *Journal of the Royal Statistical Society*, 29 (2), 119-127.
- [51] McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5 (4), 115-133.
- [52] Tilmanne, J., Urbain, J., Kothare, M. V., Wouwer, A. V., & Kothare, S. V. (2009). *Algorithms for sleep–wake identification using actigraphy: a comparative study and new results*. *Journal of sleep research*, 18(1), 85-98.
- [53] Barrat, S. (2010). *Modèles graphiques probabilistes pour la reconnaissance de formes*. (Thèse pour obtenir le grade de Docteur), l'université de Nancy 2.
- [54] Karsoliya, S. (2012). *Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture*. *International Journal of Engineering Trends and Technology*, 3 (6), 714-717.
- [55] Berry, M. J., & Linoff, G. (1997). *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc. ISBN: 978-0-471-17980-1.
- [56] John Suckling, J Parker, D Dance, S Astley, I Hutt, C Boggis, I Ricketts, E Stamatakis, N Cerneaz, S Kok, et al. (1994) The mammographic image analysis society digital mammogram database. In *Excerpta Medica. International Congress Series*, volume 1069, pages 375–378.
- [57] <http://figment.csee.usf.edu/Mammography/Database.html> [Accès le 2019 05 07].
- [58] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2000, June). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography* (pp. 212-218). Medical Physics Publishing.
- [60] Rosenblatt F., (1958). *The perceptron : a probabilistic model for information storage and organization in the brain*. *Psychological review*, 65 (6), 386-408.
- [61] Hopfield, J. J. (1982). *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the national academy of sciences*, 79 (8), 2554-2558.

[62] Kohonen T., (1989). *Self-organization and associative memory*. Springer Berlin Heidelberg. ISBN: 978-3-540-51387-2.