

Index des symboles

Symboles	Significations
IRM	Imagerie par résonance magnétique.
SAD	Système assisté par ordinateur.
Mias	Mammography Image Analysis Society
MLO	la médiolatérale oblique.
CC	l'incidence Cranio Caudale.
RNMC	Réseaux de neurones.
SVM	Support à vecteur machine.
DDSM	Digital Database for Screening Mammography.
SGLDM	La dépendance de niveaux de gris spatiale.
SRDM	La dépendance de la région environnante.
GLRM	La méthode de déroulement de niveaux de gris.
GLMD	La différence de niveau de gris.
ROI	La région d'intérêt.
RDF	La reconnaissance de forme.
RX	Rayon X.
ACR	American College of Radiology.
BIRADS	Breast Imaging Reporting and Data System.
Mcs	Microcalssification.
SE	Sensibilité.
SP	Spécificité.
Taux	Taux de classification.
FP	Faux positif.
FN	Faux négatif.
VN	Vrai négatif.
VP	Vrai positif.
NG	Niveau de gris.
RS	Rapport d'asymétrie.

Résumé

Le cancer du sein, forme la première cause de mortalité chez les femmes dans le monde, Aujourd'hui, Il est devenu une urgence pour lutter contre ce type de cancer. Pour cette raison, plusieurs études ont montré la nécessité d'une lecture attentive des clichés mammographiques pour aider les médecins dans le dépistage précoce de ce dernier.

L'objectif est de développer des systèmes puissants pour renforcer la reconnaissance des zones anormales, qui sont connus sous le nom SAD (système d'aide au diagnostic). nous développons des classifieurs pour l'identification des masses mammaires en se basant sur les techniques des RNs (réseaux de neurones) et le SVM (support à vecteur machine).

Notre but principal est de classifier les masses mammaires après la détection et la caractérisation des différents descripteurs contenus à l'intérieur de l'image mammographiques. En s'appuyant sur la matrice de cooccurrence. Nous avons utilisé les informations textuelles contenues dans les tissus mammaires exprimées en niveau de gris, et sur les descripteurs de formes géométriques. Le choix de ces descripteurs constitue un problème majeur dans les systèmes d'analyse d'images, car ces derniers conditionnent fortement le résultat final de la classification. Donc, il s'agit de trouver une caractérisation des images plus riche en information que les images d'origine et ceci dans le but de rendre la prise de décision plus efficace. Nous avons validé nos classifieurs sur un échantillon de 90 images de la base de données DDSM. Les résultats expérimentaux ont montré l'intérêt des descripteurs choisis.

Mots clés : *Cancer du sein, Diagnostic Assisté par Ordinateur (DAOx), Segmentation, classification, Descripteurs de texture et de forme, SVM, RNs.*

Abstract

Breast cancer, the leading cause of death among women worldwide, today is urgently improve to fight against this cancer .For this reason, several studies proved the nessecity of the second reading of mammographical cliches to help doctors in the early detection of the latter.

The objective is to develop powerful systems to reenforce the recognition of abnormal areas, which are known as Computer Assisted diagnosis (CAD). One of these systems will we use in the diagnosis of breast masses is based on the supervised learning technique using SVM and neural networks (RNs).Our main goal is to classify breast masses after the detection and characterization of various descriptors contained within the mammographic image, based on the co-occurrence matrix in order to encode information in textural fabrics Breast expressed in grayscale, and the descriptors of geometric shapes. The choice of these descriptors is a major problem in image analysis systems because these descriptors strongly influence the final result of the research or of the classification. So it's the best characterization of the cibled image which make final decision more effective. We have validated our system on a sample of 90 frames(DDSM) and the experimental results show the effectiveness of these descriptors.

Keys words : *Breast cancer,classification,segmentation,SVM,RNs, Feature Extraction,Computer Assisted diagnosis(CAD).*

Table des matières

Introduction générale	ix
1 Principes anatomiques	1
1.1 Introduction	1
1.2 Épidémiologie du cancer du sein en Algérie	1
1.3 Anatomie du sein	1
1.4 Définition du cancer du sein	2
1.5 les différentes tumeurs	2
1.5.1 Les tumeurs bénignes	2
1.5.2 Les tumeurs malignes	2
1.5.3 Causes du cancer du sein	3
1.5.4 Le cancer du sein chez l'homme	4
1.6 Les examens mammographiques	4
1.6.1 Dépistage	4
1.6.2 Diagnostic	4
1.6.3 Traitement	4
1.7 Les techniques d'imagerie mammographiques	5
1.7.1 La mammographie	5
1.7.1.1 La technique	5
1.7.2 Échographie mammaire	6
1.7.3 L'imagerie par résonance magnétique	6
1.7.4 Les incidences en mammographie	7
1.8 Les signes radiologiques	7
1.8.1 Les opacités du sein	7
1.8.1.1 L'analyse des opacités mammaire	8
1.8.2 Les calcifications	10
1.8.2.1 Les macro-calcifications	10
1.8.2.2 Les micro-calcifications	10
1.9 Le système SAD	11

1.9.1	La classification des pathologies mammaires.	11
1.9.1.1	La classification de Le Gal	11
1.9.1.2	La classification BIRADS	12
1.10	Conclusion	14
2	Processus automatisé De reconnaissance des Masses mammaires	15
2.1	Introduction	15
2.2	L'état de l'art	15
2.3	Présentation des bases d'images mammographiques	17
2.3.1	Digital Data base for Screening Mammography (DDSM)	17
2.3.2	La base de données (MIAS)	17
2.3.2.1	Le type de la densité mammaire	18
2.3.2.2	Le type de la lésion si elle existe	18
2.4	Approche proposée	19
2.4.1	prétraitement	19
2.4.2	Segmentation	20
2.4.2.1	Principe du FCM	20
2.4.3	L'extraction des caractéristiques	20
2.4.3.1	Caractéristique couleur	21
2.4.3.2	Caractéristique de Texture	21
2.4.3.3	Caractéristique de forme	23
2.5	Modèles de classification	23
2.5.1	La classification supervisée	23
2.5.2	La classification non supervisée	24
2.5.3	La classification semi-supervisée	24
2.6	Conclusion	24
3	Caractérisation Des Masses Mammaires	25
3.1	Introduction	25
3.2	L'état de l'art	25
3.3	L'extraction des caractéristiques	26
3.3.1	L'extraction d'attributs texturaux dans les images en niveau de gris	27
3.3.1.1	Statistiques du premier ordre	27
3.3.1.2	Statistiques du deuxième ordre	28
3.3.1.3	Matrice de cooccurrence :	29
3.4	Attributs d'Haralik	30
3.5	Les descripteurs de formes	33

3.5.1	Approche basée contour	34
3.5.2	Approche basée région	34
3.5.3	Attributs géométriques	35
3.6	Conclusion	40
4	Résultats et Discussion	41
4.1	Introduction	41
4.2	Classification neuronal :	41
4.3	Classification par SVM	42
4.3.1	Principe de l'algorithme :	42
4.4	Présentation de la base DDSM d'images mammographiques :	43
4.4.1	Le choix de la base DDSM :	43
4.5	Le choix des descripteurs :	44
4.5.1	La caractérisation de la masse mammaire :	46
4.5.2	Présentation de notre modèle de traitement :	47
4.6	Les caractéristiques de la masse :	48
4.6.1	la forme :	48
4.6.2	la texture :	50
4.7	Classification :	51
4.7.1	Base d'apprentissage :	52
4.7.2	Phase de test :	52
4.8	Les résultats obtenus :	52
4.9	Conclusion :	53
	Conclusion générale et perspectives	55
	Références bibliographiques	57

Table des figures

1.1	Anatomie interne du sein	2
1.2	Types de cancers du sein	3
1.3	Les composants d'une mammographie.	5
1.4	Exemples d'images échographiques du sein	6
1.5	Exemple d'IRM mammaire bi-sein.	6
1.6	Incidence en mammographies.	7
1.7	Radio anatomie d'un sein normale sur une incidence face/oblique.	7
1.8	Exemple des opacités bénigne /maligne.	8
1.9	Différents types de densité mammaire.	8
1.10	Les formes de masse.	9
1.11	Exemple de calcifications malignes.	9
1.12	Exemple des micro-calcifications.	10
1.13	Exemple de calcifications bénignes.	10
1.14	Classification de Le Gal.	12
2.1	Des échantillons de la base DDSM.	17
2.2	Quelques images de la base de données MIAS.	18
2.3	Diagramme général d'un système SAD.	19
2.4	Extraction du contour de la masse.	20
2.5	Information locale et information globale.	22
3.1	Extraction des caractéristiques.	27
3.2	Plus proches voisins du pixel 'X' selon les quatre directions.	29
3.3	la matrice de cooccurrence avec $D=1$ et PH (0°),PV (90°),PRD (45°),PLD (135°).	29
3.4	Classification des techniques de la représentation de forme[47].	35
3.5	Calcul de la mesure de circularité.	36
3.6	Calcul de la mesure de rectangularité.	36
3.7	l'ellipse équivalente d'un objet.(a) l'image originale ,(b) le résultat.	37
3.8	(a) Convexe et ses concavités ; (b) l'arbre de concavité [45].	38

3.9	Construction de l'axe médian d'une masse maligne.	39
3.10	Exemple montrons les propriétés du squelette.	39
3.11	Exemple montrons la continuité du squelette.	40
4.1	perceptron multicouches.	42
4.2	représente un hyperplan.	43
4.3	Représentation de l'évolution du signe de malignités des masses mam- mographiques.	44
4.4	Cas 0004 de la base DDSM avec annotations.	45
4.5	chaîne de traitement abordée.	47
4.6	Extraction de mesures morphologiques à partir du masque binaire ou de contour.	48
4.7	Le rapport d'asymétrie, la compacité et le périmètre des masses mam- maires. (a) masse maligne, (b) masse bénigne.	48
4.8	Les points terminaux du squelette.	49

Liste des tableaux

1.1	Tableau récapitulatif des opacités bénignes/malignes.	11
1.2	Tableau Classification BIRADS de l'ACR.	13
4.1	Répartition de la base utilisée.	44
4.2	Récapitulatif de mesure d'attributs de forme	49
4.3	Récapitulatif de mesure d'attributs de texture.	51
4.4	Classification par SVM.	53
4.5	Classification neuronal.	53

Introduction générale

Un cancer concerne toutes les tumeurs malignes qui s'étendent plus ou moins rapidement d'une façon très anarchique comme le cancer du sein qui est un réel enjeu de santé publique.

En Algérie, le cancer du sein représente près de 50% des cancers gynécologiques chez la femme et au cours de ces quinze dernières années, l'incidence du cancer du sein a été multipliée par trois. Du fait de son diagnostic tardif, il en résulte souvent un traitement lourd, mutilant et coûteux qui s'accompagne d'un taux de mortalité élevé [1].

Malgré les progrès récents, la problématique qui reste toujours posée concerne la complexité de l'étape de détection de ces lésions (l'aspect très texturé des tissus mammaires, la diversité de la taille des masses, la faible intensité entre les tumeurs et les tissus sains qui rendent les bords de ces masses ambigus et flous). Le but majeur et d'augmenter la sensibilité de la mammographie afin de réduire le nombre de faux négatifs (FNs). La mammographie reste la technique d'imagerie la plus fiable et la plus reproductible pour le diagnostic précoce du cancer du sein, l'interprétation et l'analyse de ce type d'images restent une tâche plus difficile d'où le besoin à des systèmes d'aide au diagnostic pour aider les radiologues, à faire la bonne lecture des images mammographiques.

La deuxième étape aussi intégrante dans un processus de diagnostic, touche le volet de la classification sachant que la diversité des causes de malignité comme : contour, forme, densité, distribution, ...). Pour cela, des études ont montré que les SADs améliore la sensibilité de plus de 15% [2]. Elle constitue une approche standardisée d'interprétation basée sur une deuxième lecture des images mammographiques. Dans ce cadre, plusieurs travaux ont été menés ces dernières années afin de développer des systèmes d'aide au diagnostic (SAD) qui a pour objectif d'obtenir un taux maximum de reconnaissance des zones anormales.

Ce système repose sur un schéma assez générale de quatre parties : pré-traitement ,la

détection des lésion mammaires, l'extraction des caractéristiques et la classification pour la prise de décision :

- La segmentation : elle consiste à extraire les zones d'intérêts à partir des images mammographiques, ce sont les masses bénignes / malignes pour notre application.
- La caractérisation : elle fournit des caractéristiques des objets précédemment extraits. Ces caractéristiques doivent être précises pour décrire les zones d'intérêts.
- La classification : les zones d'intérêts sont classées selon le type de pathologie grâce à une méthode de classification.

L'objectif visé dans ce mémoire consiste à développer une approche de traitement et d'analyse d'images (les étapes de segmentation et de caractérisation), ce qui apportera une aide précise au diagnostic. Dans ce travail nous nous intéressons aux descripteurs de la forme et de la texture afin de classer les pixels des lésions, en utilisant deux types de classifieurs (SVM et RNs) , Nous avons réalisé dans le cadre de ce mémoire de fin d'étude les objectifs suivante :

- Détecter les zones d'intérêts en utilisant l'algorithme *FCM (Fuzzy C-Means)* ;
- La caractérisation des images (texture/ forme).
- Analyser les textures, les paramètres de forme géométriques des zones d'intérêts pour classifier les masses.
- Classification des images mammographiques.

Plan et contenu de mémoire

Le premier chapitre présente une description courte de l'anatomie du sein, le cancer du sein. Nous décrivons ensuite, quelques notions générales sur les différents types d'imagerie mammaire et nous détaillons ensuite, les signes radiologiques. Nous terminons ce chapitre par une présentation de la classification des lésions mammaires selon LEGAL et BIRARDS. Dans le deuxième chapitre, nous présentons l'état de l'art concernant le système SAD et ses différentes étapes. Nous terminons ce chapitre par une présentation des descripteurs de la couleur, la forme et de la texture. Le troisième chapitre présente le cœur de notre travail, il est dédié à la caractérisation des masses mammaires de la texture (attributs d'Haralik) et de la forme (parmi les descripteurs géométriques : la surface, le périmètre, la compacité.) appliquée sur des images mammographiques en niveau de gris. Dans le dernier chapitre : Nous analysons les résultats obtenus par les deux types de classifieurs (RNs, SVM).

Chapitre 1

Principes anatomiques

1.1 Introduction

Le sein est un organe vital du corps humain qui est caractérisé par son organisation anatomique et fonctionnelle. Ce type de cancer constitue à travers le monde la cause de décès la plus fréquente chez la femme.

Dans ce chapitre, nous allons présenter tout d'abord une vue générale sur l'anatomie du sein, ainsi que les statistiques du cancer du sein. Puis, nous étudierons les principaux changements de la pathologie qui peuvent se produire, puis nous aborderons l'approche médicale de l'imagerie du sein, Pour finir, nous verrons des méthodes de la classification des lésions mammaires.

1.2 Épidémiologie du cancer du sein en Algérie

Il est aujourd'hui urgent de mieux lutter contre le cancer du sein qui représente la première cause la plus fréquente de décès par cancer chez les femmes (522000 décès). Une femme sur 8 développe un cancer du sein au cours de sa vie, il touche des femmes âgées plus de 50 ans.

En *Algérie*, les centres de santé du pays enregistrent entre 7000 et 10.000 nouveaux cas (mars 2014) de cancer du sein par an (de la part de CHU de *Tlemcen*).

1.3 Anatomie du sein

Le sein est essentiellement constitué d'un tissu adipeux graisseux plus ou moins important qui lui donne sa forme et son volume. La glande mammaire, noyée dans le tissu conjonctif, est composée d'une vingtaine de lobules qui deviennent actives en période de lactation. Le lait sécrété est diverse par des canaux galactophores séparés

au niveau du mamelon. Le sein parcouru par une multitude de vaisseaux sanguins, est maintenu par la peau qui le recouvre et par des fibres liées au muscle pectoral [3]. La figure 1.1 détaille ces différents composants mammaires.

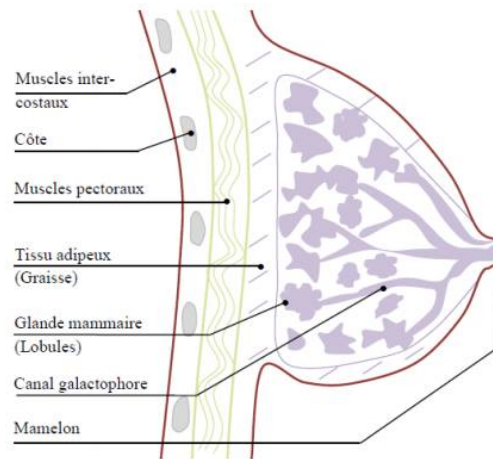


FIGURE 1.1 – Anatomie interne du sein

1.4 Définition du cancer du sein

Un cancer du sein est une lésion cellulaire qui se développe au niveau des glandes mammaires ou des canaux galactophores (conduits qui transportent le lait jusqu'au mamelon) ou bien d'une cellule des lobules du sein responsables de la production du lait (carcinome lobulaire du sein). Cependant, lorsque la cellule devient cancéreuse, elle perd ses capacités de réparation. Elle se met alors à se multiplier et finit par la formation d'une masse qu'on appelle tumeur maligne (la métastase).

1.5 les différentes tumeurs

1.5.1 Les tumeurs bénignes

Les tumeurs bénignes sont des masses tumorales qui présentent un caractère régulière à croissance lente restant toujours au niveau du tissu dans lequel elles ont pris naissance et se multiplient localement. Ces types de tumeurs ne métastasent jamais.

1.5.2 Les tumeurs malignes

Il y a une grande différence entre les tumeurs bénignes et malignes, les tumeurs malignes se croient rapidement d'une façon anarchique, et se disséminent à distance

par voie sanguine ou lymphatique en donnant des tumeurs secondaires dans d'autres viscères c'est ce qu'on appelle les métastases. L'évolution de ces tumeurs sans traitement tendre spontanément vers la mort. Ces tumeurs sont mal limitée, leur forme sont irréguliers (spiculé)(voir figure 1.2).

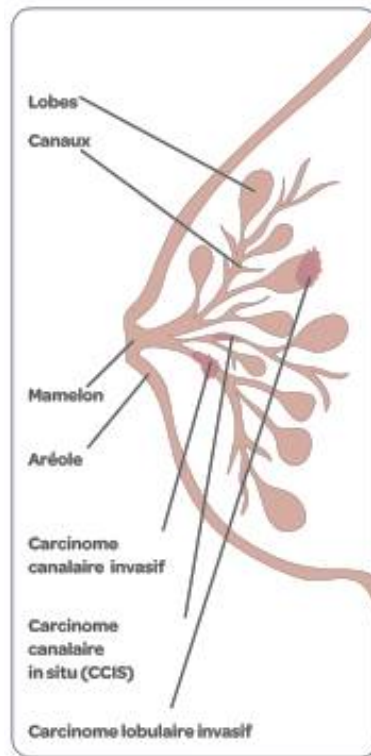


FIGURE 1.2 – Types de cancers du sein

1.5.3 Causes du cancer du sein

Il existe plusieurs causes responsables du cancer, On cite :

- L'âge : le cancer du sein affecte les femmes de plus de 50 ans.
- Les antécédents familiaux.
- Le tabagisme.
- Des seins denses.
- la radiothérapie au thorax
- L'obésité.

1.5.4 Le cancer du sein chez l'homme

Le cancer du sien est rare chez les hommes. Ils possèdent des seins qui sont toutefois moins développés que chez les femmes. Il est important que les hommes sachent qu'ils peuvent être concernés par ce type de cancer.

1.6 Les examens mammographiques

1.6.1 Dépistage

Aujourd'hui, l'examen de mammographie reste 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 (dues à la tumeur ou à la structure mammaire environnante) ou à des facteurs liés au lecteur, la distraction ou le niveau d'expérience du radiologue. Les lésions manquées sont très souvent liées à un manque 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. Le taux de faux négatifs en première lecture est loin d'être négligeable puisqu'il a été évalué entre 16 et 31%. La double lecture réduit le taux de faux négatif de 5 à 15%, en améliorant particulièrement la détection des cancers de petite taille. 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 dépistage de masse, mais elle est de plus en plus difficile à organiser aujourd'hui, en raison du nombre insuffisant de radiologues spécialisés dans la lecture des mammographies [4].

1.6.2 Diagnostic

Le diagnostic du cancer du sein repose sur :

- Diagnostic clinique
- Mammographie
- Anatomopathologie.

1.6.3 Traitement

Les méthodes de traitement sont :

- Chirurgie.
- Radiothérapie.

- Chimiothérapie.
- Hormonothérapie.
- Thérapie biologique.

1.7 Les techniques d'imagerie mammographiques

1.7.1 La mammographie

technique d'imagerie médicale qui sert à détecter les anomalies mammaires avant même qu'elles n'aient provoqué des symptômes cliniques.

1.7.1.1 La technique

L'appareil dédié à la réalisation d'une mammographie est la mammographie (voir figure 1.3). Cet appareil se compose d'un tube radiogène générateur de rayons X de faible énergie (entre 20 et 50 μm) et d'un système de compression du sein. En premier temps, les deux seins sont comprimés à tour de rôle. Cette compression permet l'étalement des tissus mammaires ce qui facilite la visualisation des structures du sein et la réduction de la dose de rayons X délivrée. En deuxième temps, les deux seins sont exposés à une faible dose de rayons X. On obtient, alors, une projection du sein sur un détecteur plan. La radiographie est réalisée sur des films argentiques ou sur des systèmes de radiologie digitale de haute qualité. L'analyse de la glande mammaire est réalisée grâce aux différences de l'atténuation des différents types de tissu [3].

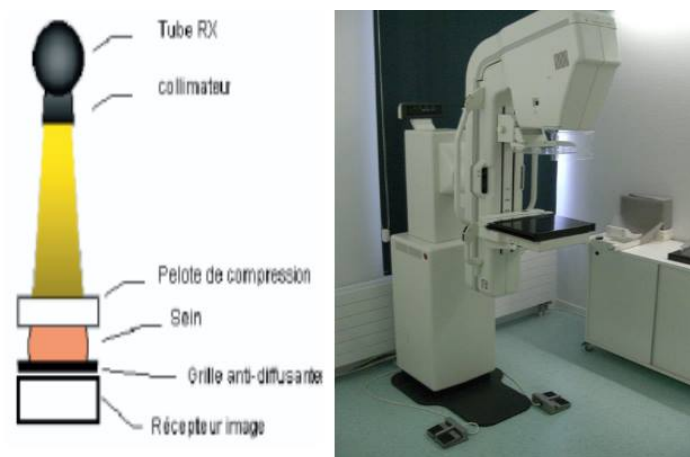


FIGURE 1.3 – Les composants d'une mammographie.

1.7.2 Échographie mammaire

Le but de cet examen est de visualiser vos seins à l'aide d'ultrasons. Il est sans danger et indolore. L'échographie du sein apporte des informations complémentaires à la mammographie. Elle est toujours utilisée chez les patientes jeunes et pour les femmes dont les seins sont denses à la mammographie. Elle sert aussi à guider le radiologue pour effectuer un prélèvement (biopsie) ou pour poser un repère métallique avant une chirurgie (voir figure 1.4)[5].

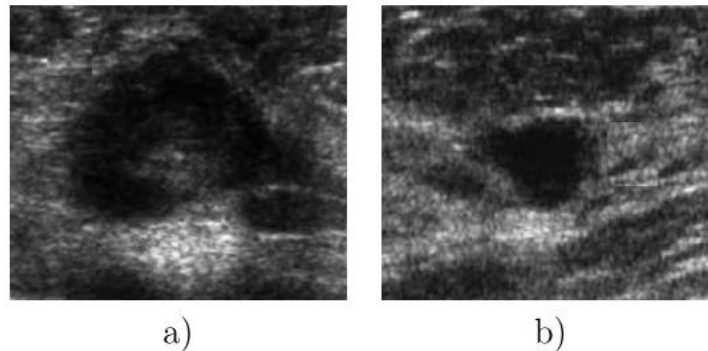


FIGURE 1.4 – Exemples d'images échographiques du sein

1.7.3 L'imagerie par résonance magnétique

L'imagerie par résonance magnétique du sein est un examen de deuxième intention doit être réalisé après un bilan sinologique complet (examen clinique, mammographie associée ou non à une échographie). Elle est beaucoup plus performante pour la détection des carcinomes infiltrant (sensibilité allant de 94% à 100%) que pour celle des carcinomes intracanalaires (sensibilité variant de 77% à 94%) [6](voir figure 1.5).

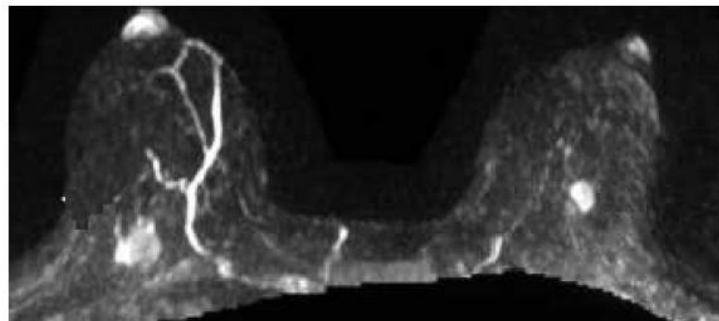


FIGURE 1.5 – Exemple d'IRM mammaire bi-sein.

1.7.4 Les incidences en mammographie

La zone d'intérêt vue sur deux incidences différents les plus courantes est l'incidence oblique externe (ou medio-latéral-oblique) et l'incidence de face (voir figure 1.6).



FIGURE 1.6 – Incidence en mammographies.

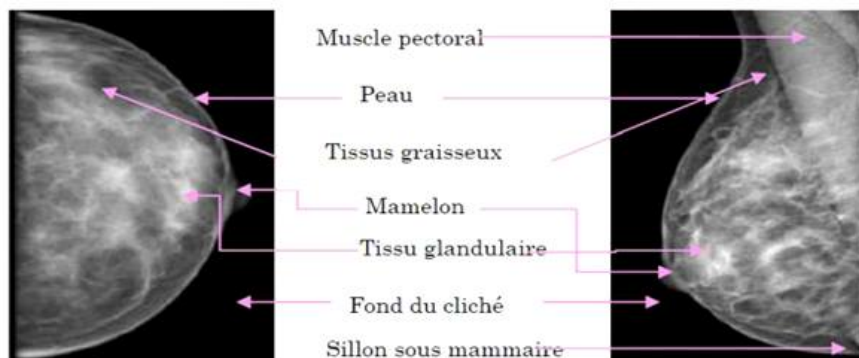


FIGURE 1.7 – Radio anatomie d'un sein normale sur une incidence face/oblique.

1.8 Les signes radiologiques

1.8.1 Les opacités du sein

Une opacité est une masse importante correspond à une plage de sur-densité anormale. Une sur-densité anormale ne se distingue pas d'une sur-densité normale par un critère précis, mais par une combinaison de différentes caractéristiques : taille, densité, contour, forme, texture vue sur deux incidences différentes [7](voir figure 1.8).

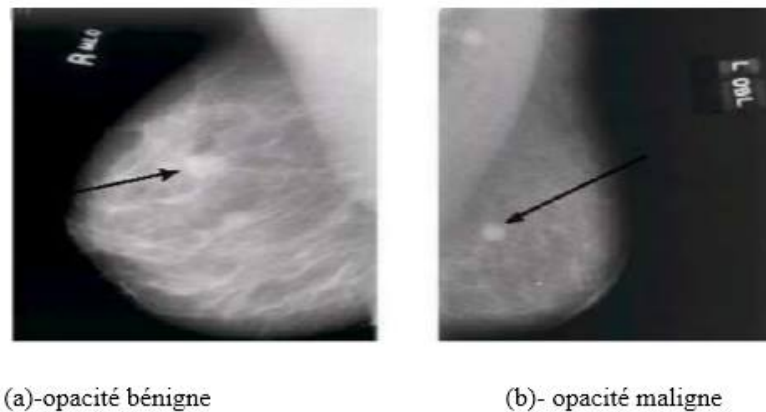


FIGURE 1.8 – Exemple des opacités bénigne /maligne.

1.8.1.1 L'analyse des opacités mammaire

Il existe quatre critères radiologiques :

- La taille : elle est variable de quelque millimètre à plusieurs centimètres. Cependant la taille ne prédit pas le caractère malin, sauf sur des clichés successifs lorsque l'on voit la taille augmenter régulièrement.
- La densité : Les masses sont généralement dense (plus blanches) que le tissu mammaire adjacent. Il existe quatre types (voir figure 1.9) :

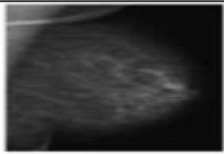
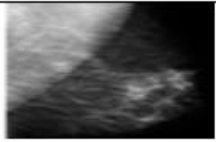

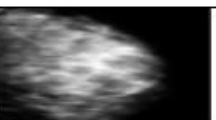
Types	Exemples	Aspect morphologique
Type 1		Les seins entièrement clairs, adipeux (graisseux).
Type 2		Structure fibroglandulaire peu importante.
Type 3		dense de façon hétérogène avec difficultés d'interprétation.
Type 4		Sein uniformément dense avec difficultés d'interprétation.

FIGURE 1.9 – Différents types de densité mammaire.

- La forme : Cinq types de formes (voir figure 1.10) de masse :

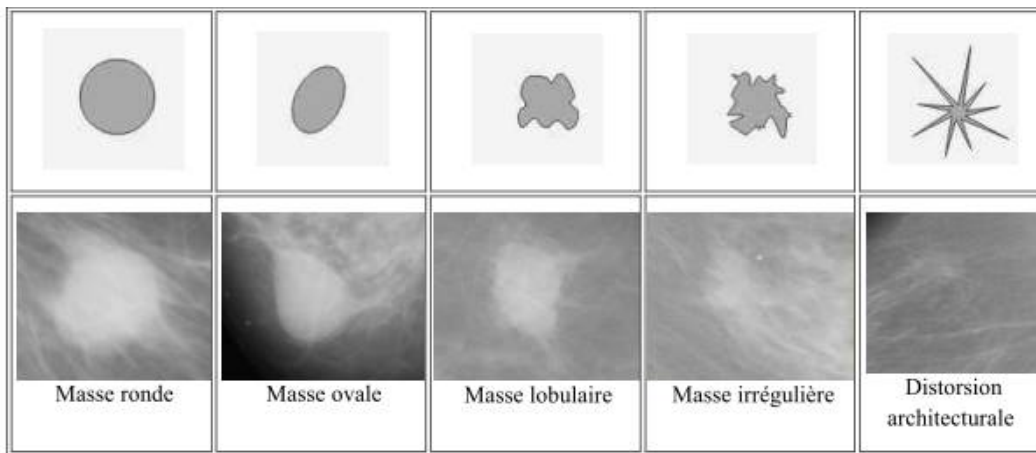


FIGURE 1.10 – Les formes de masse.

- Le contour : c'est le meilleur critère pour l'étude des opacités [7]. La figure 1.11 illustre cinq types de contours des masses :

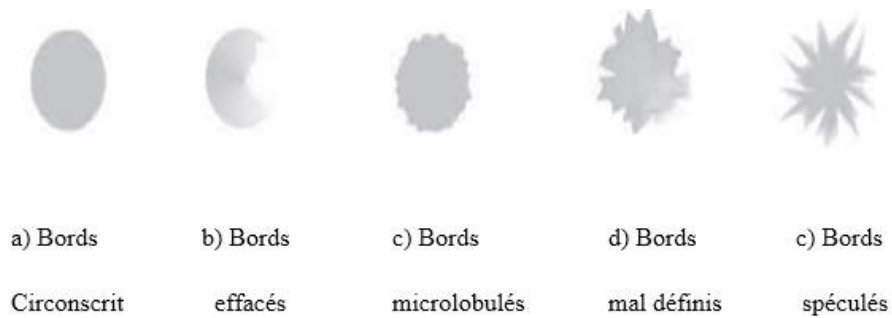


Figure 11 Les contours de masse.

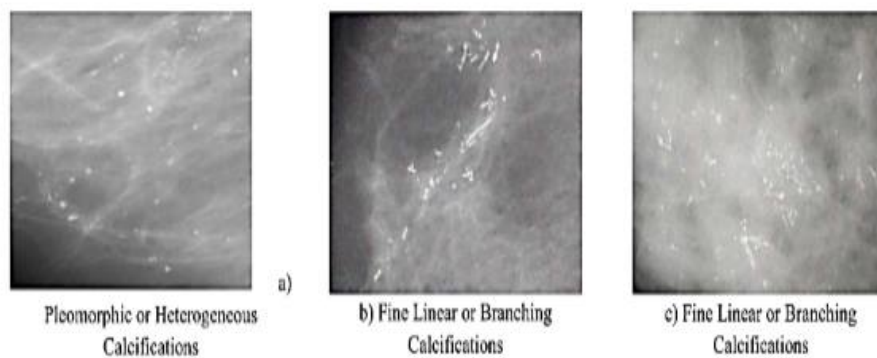


FIGURE 1.11 – Exemple de calcifications malignes.

- Thérapie biologique.

1.8.2 Les calcifications

Il existe deux types de calcifications : les macro-calcifications et les micro-calcifications

1.8.2.1 Les macro-calcifications

Sont des dépôts grossiers de calcium blanc, opaque aux RX et bien visible en mammographie. Elles sont souvent associées à des signes bénins.

1.8.2.2 Les micro-calcifications

Sont des calcifications de petite taille en général comprises entre 200 et 500 μm (voir figure 1.12). L'analyse des calcifications se fait selon leur taille, forme, nombre et

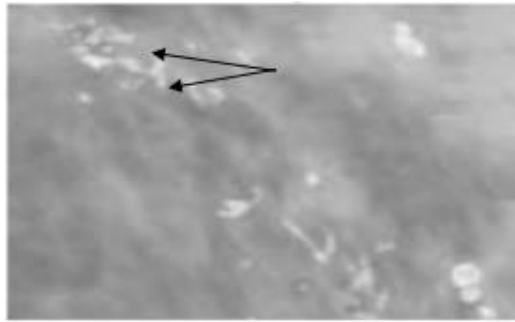


FIGURE 1.12 – Exemple des micro-calcifications.

leur distribution. En règle générale, les calcifications de grande taille régulière, ronde ou ovale sont synonymes de tumeur bénigne, les calcifications petites, irrégulières, de formes variées sont souvent associées avec une pathologie maligne (figure 1.13).

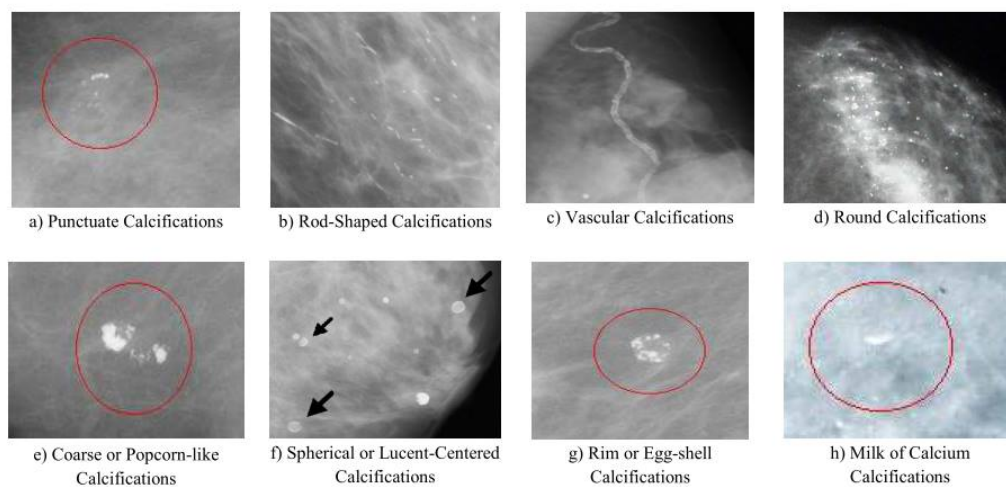


FIGURE 1.13 – Exemple de calcifications bénignes.

Paramètres radiologiques	Opacité positive	Opacité négative
Taille	-Les micro-calcifications sont associées avec un processus malin. - Moins de 0.5 mm sont souvent associées à des cancers.	-Les macro-calcifications avec un processus bénin. -Celles de plus de 2 mm sont assez typiques de lésions bénignes.
Forme	-zone de flou limité	-parfaitement net
Morphologie	-polylobée, avec des rayons de courbure variée	-ronde ou ovales, sans lobulation ni rectitude localisée

TABLE 1.1 – Tableau récapitulatif des opacités bénignes/malignes.

1.9 Le système SAD

Depuis une quinzaine d'années, plusieurs équipes se sont intéressées à des systèmes informatiques automatisés pour l'aide à la détection et au diagnostic des lésions radiologiques en mammographie. L'analyse numérique et informatique des clichés de mammographie peut alors servir de deuxième lecture informatique pour le radiologue. Le but du système expert est d'augmenter la sensibilité de la mammographie tout en réduisant le nombre de faux positifs. Plusieurs études récentes ont montré que ces systèmes SAD pouvaient augmenter de 15 à 20 % le taux de cancers détectés en situation de dépistage. Il a également été montré que les techniques d'intelligence artificielle pouvaient mieux classer les anomalies détectées en lésions bénignes ou malignes qu'un radiologue non spécialisé en mammographie [4].

1.9.1 La classification des pathologies mammaires.

L'étude morphologique des lésions mammaires a fait l'objet de plusieurs classifications, les plus connues et les plus pratiquées sont celles de Le Gal et de BIRADS [3].

1.9.1.1 La classification de Le Gal

Cette classification cherche à évaluer le caractère dangereux ou non des calcifications observées. Toutefois, son défaut principal est qu'elle se base uniquement sur les micro-calcifications et n'intègre pas d'autres paramètres tels que :

- L'étude des distorsions architecturales.
- Le comportement du radiologue vis-à-vis de chaque cas.

- L'étude morphologique des masses.
- La disposition des micro-calcifications[3].

Elle décrit cinq types morphologiques qui ont une valeur prédictive de malignité croissante (figure 1.14) :

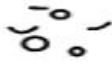




Type 1 : 	Mcs annulaire, arciforme ou polyédrique. Risque de cancer du sein quasi nul (0% de cancers).
Type 2 : 	Mcs rondes et de taille variables. Risque de carcinome : 22%.
Type 3 : 	Mcs poussiéreuse, pulvérulente. Risque de cancer : 36%.
Type 4 : 	Mcs irrégulière associée à un risque de cancer : 56%.
Type 5 : 	Mcs vermiculaire ou branchée. Risque de carcinome : 90%.

FIGURE 1.14 – Classification de Le Gal.

1.9.1.2 La classification BIRADS

Pour palier aux lacunes de la classification de Le Gal, l'American College of Radiology (ACR) a souligné l'importance d'un protocole mammographique standardisé et complet qui tient en compte les différents facteurs de malignité [3]. En novembre 1998, l'ACR a établi le système Américain BIRADS (Breast Imaging Reporting and Data System) qui a été rédigé par un groupe d'experts réunis par l'ANAES 2 . Ce système permet de classer les images mammographiques en six catégories en fonction du degré de suspicion de leur caractère pathologique (tableau 1.2).

Type	Aspects mammographiques	Interprétation
ACR		
ACR0		Cliché insuffisant pour vérifier la présence ou non de lésions.
ACR1	Mcs rondes à centre clair radio-transparentes.	Absence de masses suspectes et de micro-calcifications suspects.
ACR2	Mcs punctiforme, rondes, pleines, radio-opaques, aux contours réguliers et arrondis.	Présence d'opacité ronde avec macro-calcifications sans opacité, Images de densité graisseuse ou mixte. Les Opacités rondes correspondant à un kyste en échographie. Présences des Calcifications punctiformes régulières diffuses.
ACR3	Mcs trop fine pour préciser leur forme elles donnent une image de semis de poudre calcaire.	Masse circonscrite ronde ou ovale sans microlobulations non liquidienne en échographie. Mcs ronde ou punctiforme régulière ou pulvérulente, peu nombreuse, en petit amas rond isolé. Petit amas rond ou ovale de calcifications amorphes et peu nombreux.
ACR4	Mcs punctiforme irrégulière aux contours anguleux, différents d'une micro-calcification à l'autre.	Masse spiculée sans centre dense, ronde ou ovale à contour lobulé ou masqué. Mcs pulvérulente groupée et nombreuse, irrégulière, polymorphe ou granulaire, peu nombreuse.
ACR5	Mcs vermiculaire, elles ont la forme d'un bâtonnet souvent	Masse spiculée à centre dense, à contour flou et irrégulier. Mcs associée à une anomalie architecturale ou à une opacité.

TABLE 1.2 – Tableau Classification BIRADS de l'ACR.

1.10 Conclusion

Dans ce chapitre, nous avons décrit l'anatomie du sein ainsi que les différentes modalités d'imagerie mammaire. Le deuxième chapitre est consacré à détecter les masses mammaires par des outils d'aide au diagnostic et pour donner une deuxième lecture aux médecins. Pour cela, nous avons besoin d'une connaissance suffisante sur les caractéristiques des masses mammaires qui vont être présentées par la suite.

Chapitre 2

Processus automatisé De reconnaissance des Masses mammaires

2.1 Introduction

Dans le cadre de mémoire de master. Nous avons développés trois modules distincts, le premier module concerne la phase de segmentation, le deuxième module permet la caractérisation des images mammaires via des descripteurs de texture et de forme.

Enfin dans le dernier module, nous implémentons deux classifieurs différentes (RNs, SVM) dans la littérature nous trouvons beaucoup de travaux qui concernent la segmentation, la caractérisation et la classification des masses mammaires.

2.2 L'état de l'art

Le cancer du sein est le plus fréquent chez les femmes à travers le monde il est classé parmi les causes principales de décès. C'est une lésion maligne qui se développe d'une façon anarchique au niveau de la glande mammaire (adénocarcinome). Actuellement, Il n'existe pas un moyen précis pour éviter son apparition. Les radiologues spécialistes ne détectent que 70% de cas de cancer du sein à cause de la difficulté de l'examen mammographique. Pour cette raison, plusieurs recherches ont été proposées ces dernières années afin de développer des outils d'aide au diagnostic (SAD) dont l'objectif est d'interpréter les images de mammographies. Ces outils présentent une chaîne compacte. Après la segmentation des images mammographiques, l'étape suivante est la caractérisation de l'image ensuite la classification afin d'identifier le type de tumeurs.

- Une phase de pré-traitement qui suit l'acquisition et la numérisation de l'image est utilisée pour éliminer les bruits. [8] proposent une transformation des niveaux de gris en fonction d'éléments locaux comme les contours ou les statistiques locales du premier ordre. D'autres méthodes ont aussi été proposées comme l'utilisation de filtre d'amélioration de la netteté [8], ou encore la suppression du fond de l'image [9].
- Une phase de segmentation permet d'isoler les régions d'intérêts qui peuvent contenir des masses. Certains chercheurs [10] ont utilisé un nouveau type d'algorithme de seuillage (minimisation de la somme de l'énergie) pour segmenter les régions de masses où la valeur du seuil convenable est déterminée après plusieurs itérations. Aussi [11] ont fait appel à une mesure de détection des sur-densités en calculant la proportion de pixels situés autour d'une lésion qui ont une intensité inférieure au minimum de l'intensité à l'intérieur. [12] a proposé l'utilisation d'outils de morphologie mathématique pour l'extraction automatique d'opacités. [13] a proposé l'utilisation de filtre en quadrature pour la détection de ces spicules sur le fait que les lésions malignes présentent souvent une forme irrégulière (des spicules). Aussi [14] a présenté une nouvelle approche pour détecter les micro-calcifications. Ils ont utilisé l'algorithme de segmentation la ligne de partage des eaux basé sur le gradient morphologique multiéchelle.
- L'avant-dernière étape est l'extraction des caractéristiques liées à des lésions (les micro-calcifications et les masses. En particulier, nous citons principalement la texture qui est l'une des caractéristiques les plus importantes utilisées pour identifier un objet dans une image. La dépendance de niveaux de gris spatiale (SGLDM) et la dépendance de la région environnante (SRDM), la méthode de déroulement de niveaux de gris (GLRM) et la différence de niveau de gris (GLMD) sont utilisées pour extraire les caractéristiques de l'image segmentée [15]. Dans ce contexte ; les auteurs [16] ont utilisé la technique de transformation en ondelettes pour extraire les paramètres de texture de la région d'intérêt (ROI).
- la détection des lésions mammaire est considérée comme un problème de classification. Il y a plusieurs travaux de classification sur les images issues d'une mammographie, dans la littérature. [17] Ont utilisé les RNs pour la classification des mammographies et ont obtenu un taux de classification de 97.3%. Dans l'article [18], les auteurs ont utilisé une méthode basée sur l'augmentation du contraste de densité pondérée et ont obtenu un taux de classification de 82.33%. Christoyianni et al. [19] Ont utilisé la méthode d'analyse de composantes indépendantes et ont obtenu un taux de 79.31% pour la reconnaissance des lésions malignes et bénignes. Nous pouvons citer les travaux de Brook et al concernant les travaux sur les images

de biopsie. Dans [20], les auteurs ont utilisé le classifieur SVM pour classer des images segmentées par la méthode du seuillage adaptatif et ont obtenu un taux de classification de 95.37%. Une approche utilise l'algorithme Support de Vecteur Machine (SVM) pour la détection des masses mammographiques qui ont démontré leur efficacité dans le diagnostic du cancer du sein [21].

2.3 Présentation des bases d'images mammographiques

2.3.1 Digital Data base for Screening Mammography (DDSM)

La base de données Marathon de l'université de la Floride du Sud. Une description de cette base a été effectuée par American college of Radiolog dans le lexique de BIRADS (Breast Imaging Reporting and Data System). La base correspondante contient 2620 dossiers de patientes classées en trois groupes : normaux, bénins et malins. Chaque cas contient les vues obliques externes (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 quand ils existent(voir figure 2.1)[22].

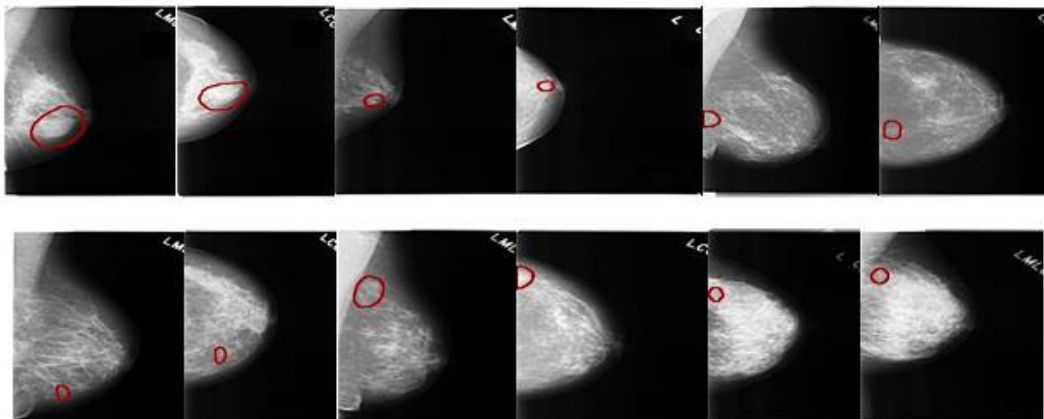


FIGURE 2.1 – Des échantillons de la base DDSM.

2.3.2 La base de données (MIAS)

MIAS : Mammographie image analysis society [23] contient un ensemble de 322 images mammographiques numérisées de type MLO (incidence latérale oblique) qui explore le sein gauche et droit de 161 femmes. Les caractéristiques de la base MIAS : les films mammographiques ont été extraits de : National Breast Screening Programme, numérisés avec une résolution de 200 μm par pixel. La base contient aussi la localisation de différentes anomalies présentes dans une mammographie, par des radiologues

experts. Marquées par un cercle de rayon donné ; qui enferme les éventuelles lésions (Foyer de (Mcs), Asymétrie, Distorsion architecturale, Masses circonscrites et spéculé, et autres anomalies). Les mammographies de format (.pgm) sont nommées : « mdb XXX », dont XXX représente le numéro de l'image qui varie de 001 au 322. La taille de chaque image est 1024 Pixels x 1024 Pixels, de NG qui varie entre le 0-255. Pour chaque mammographie, la base de données MIAS fournit des informations sur :

2.3.2.1 Le type de la densité mammaire

- F :Fatty(graisseux).
- G :Fatty glandulaire(Glandules graisseuses).
- D : Dense-glandular(Dense).

2.3.2.2 Le type de la lésion si elle existe

- CALC : Calcification (Calcification).
- SPIC :Spiculated masses.
- MISC :Other, ill-defined masses).
- ARCH :Architecturaldistortion.
- ASYM :Asymmetry.
- NORM :Normal.

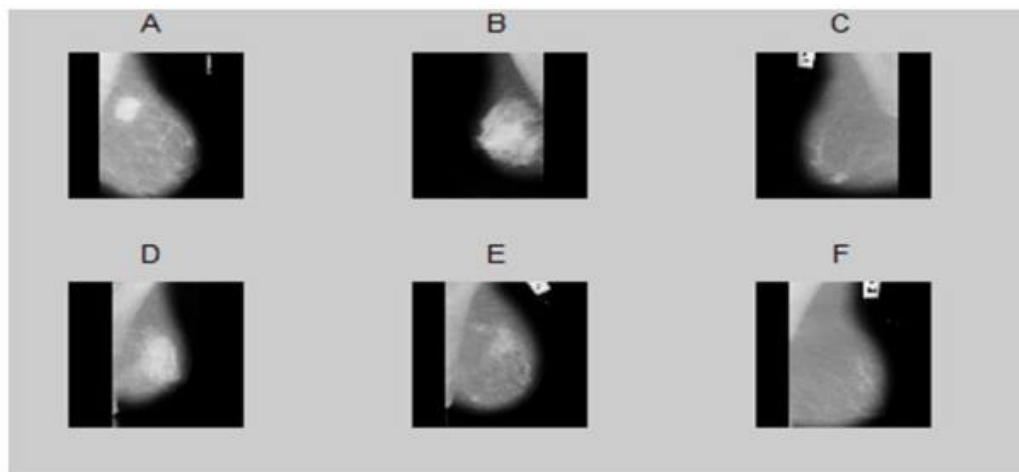


FIGURE 2.2 – Quelques images de la base de données MIAS.

2.4 Approche proposée

Notre modèle pour identifier les tumeurs est composée réellement de quatre étapes, dans ce travail nous avons gardé uniquement trois étapes sachant que les zones d'intérêt sont déjà localisées par des experts(voir figure 2.3).

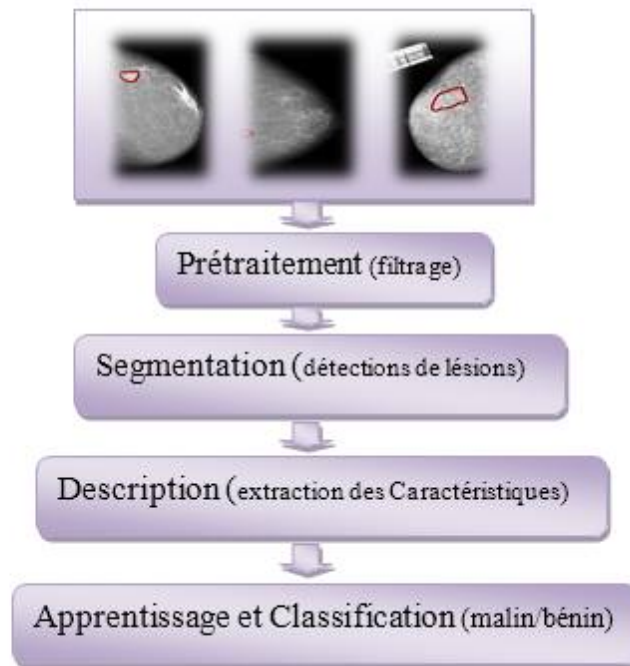


FIGURE 2.3 – Diagramme général d'un système SAD.

2.4.1 prétraitement

L'analyse d'images pour la détection des anomalies mammaires (masses /microcalcifications) débute toujours par une étape d'amélioration de contraste qui est connue sous le nom de rehaussement entre la lésion mammaire et le reste de l'image. Afin de faciliter l'interprétation visuelle ou même la classification automatique des cibles et des structures entièrement par ordinateur, l'absence de cette phase indique que plusieurs détails concernant la lésion peuvent être négligés et par la suite non diagnostiquée qui peut causer des faux négatifs (FN).ce qui ne répond pas à l'objectif principal de la détection précoce d'un cancer. Un excès de renforcement du contraste peut causer des faux positifs(FP).

Dans des précédents travaux[24], les images que nous avons utilisées ont été déjà traitées par les techniques de filtrage (élimination du bruit) et seuillage (délimitation des zones d'intérêt).

2.4.2 Segmentation

La segmentation est une phase très importante qui consiste à séparer l'image en zones ayant des caractéristiques similaires puisque les traitements ultérieurs : description (caractérisation) et classification (minimisant le taux des faux positifs(FP) et maximisant le taux des vrais négatifs(VN)) sont fortement liés au résultat de segmentation. La morphologie des masses peut être masquée partiellement par le tissu mammaire. Notre base d'apprentissage contient des images mammographiques dont la masse est entourée par un cercle rouge ,le choix de cette dernière a été très important, en raison du fait qu'il était nécessaire de trouver une méthode qui parviendrait à éliminer le contour élaboré par le radiologue. Notre choix s'est porté sur l'algorithme « Fuzzy C-Means FCM », il s'est avéré d'une grande précision dans la segmentation des masses mammaires[24].

2.4.2.1 Principe du FCM

c'est un algorithme de classification non supervisé, il introduit la notion d'ensemble flou dans la définition des classes, chaque point de l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leurs centres de gravité. Cet algorithme nécessite la connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective. La Figure illustre le résultat de la segmentation par l'algorithme FCM (voir figure 2.4).

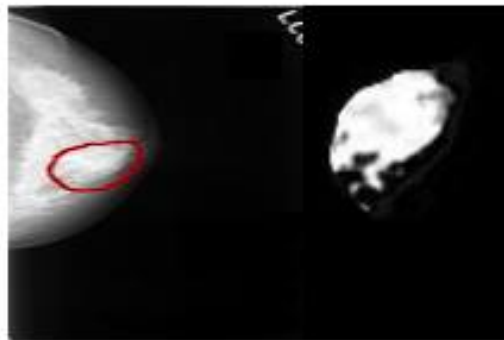


FIGURE 2.4 – Extraction du contour de la masse.

2.4.3 L'extraction des caractéristiques

En termes de malignité ou de bénignité, l'extraction des caractéristiques a pour objectif de spécifier la gravité des tumeurs. Le contenu de l'image visuelle ou bien

les descripteurs bas niveaux peuvent être divisés en trois sous-espaces : la texture, la forme et la couleur qui sont des connaissances utilisées pour caractériser l'information contenue dans les images.

2.4.3.1 Caractéristique couleur

C'est une caractéristique riche d'information et une étape qui permet aux radiologues de préciser les aspects pathologiques des anomalies mammaires. Les clichés du sein issus à partir d'une mammographie sont des images en niveau de gris qui ne possède aucune information couleur. Dans ce cas-là, ce descripteur d'images ne s'avère pas pris en compte.

2.4.3.2 Caractéristique de Texture

Dans un premier temps, Il est important de trouver les descripteurs convenables et pertinents qui permettent l'identification et la caractérisation des lésions mammaires des images mammographiques.

La texture a suscité de nombreuses recherches et travaux d'application qui ont engendré une multitude de méthodes d'analyse. Ces méthodes ont pour objet de caractériser ou décrire, de discriminer, de segmenter ou de synthétiser les textures. On ne peut dire qu'il existe une définition mathématique précise, qui est généralement trop détaillée [25]. Donc, nous citons quelques notions trouvées dans la littérature :

- [26] a défini la texture comme suit : « Une texture est une région d'une image pour laquelle il est possible de définir une fenêtre de dimensions minimales, telle qu'une observation au travers de celle-ci se traduit par une perception (impression) visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée ».
- Maitre et al [27] : « Une texture est un champ de l'image qui apparaît comme un domaine cohérent et homogène, c'est-à-dire formant un tout pour un observateur »
- La définition de Bonnevey [28] : « la texture est une combinaison de deux types d'information :
 - une information locale autour de chaque pixel (statistiques, géométriques, sémantiques)
 - une information globale qui correspond à la distribution spatiale de l'information locale (fenêtre de visualisation) »(voir figure 2.5) ;

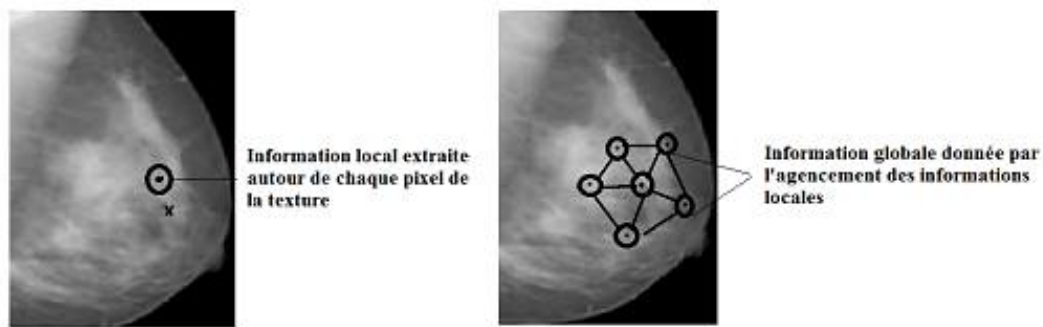


FIGURE 2.5 – Information locale et information globale.

- Nous utiliserons comme définition de texture celle de Haralick : « Une texture est décrite par un ensemble de base de primitives (ou motifs) ainsi que par leur organisation spatiale. » [29].
- On peut définir une texture en faisant une correspondance avec des adjectifs communément utilisés par les êtres humains pour décrire des textures [2]. On cite, parmi ces caractéristiques
 - La rugosité.
 - Le contraste.
 - La granularité.
 - La régularité.
 - Le volume.
 - La directionnalité.

L'information apportée par les textures des tissus du sein constitue un critère important sur lequel le radiologue se base pour analyser un mammogramme et différencier les tissus normaux des tissus pathologiques [30]. En raison, de l'inhomogénéité des tissus mammographiques qui sont caractérisés par des textures aléatoires. L'analyse statistique reste le bon choix pour ce type d'image. Parmi les outils statistiques, ils existent ceux d'ordre deux qui sont les plus utilisées pour la caractérisation de textures. La matrice de cooccurrence est la plus connue par ce qu'elle est bien adapté à la caractérisation visuelle des propriétés texturales.

La littérature est très riche à ce sujet, il existe plusieurs méthodes qui permettent d'extraire des paramètres caractérisant la texture. Elles se classent en deux grandes classes l'approche structurale consiste à repérer les motifs de base de la texture et l'approche statistique s'applique en particulier aux textures aléatoires[31].

2.4.3.3 Caractéristique de forme

L'importance de l'information locale aux régions composant l'image a fait l'objet de plusieurs chercheurs. La forme est un descripteur de bas niveau le plus performant pour caractériser les structures géométriques du contenu visuel de l'image. Par contre aux descripteurs de texture et de couleur, qui sont utilisés pour décrire les informations d'une manière précise et globale de l'image.

2.5 Modèles de classification

La classification est une technique centrale de l'étape de fouille de données dans le processus d'extraction de caractérisation dans les bases de données. Le terme classification en français désigne à la fois les termes anglais classification (classification supervisée) et clustering (classification non supervisée). L'objectif de la classification est de faire la correspondance entre les classes spectrales et les classes d'information c'est-à-dire de déterminer pour chaque pixel de l'image, sa classe d'appartenance [32]. Il existe plusieurs méthodes de la classification, nous citons parmi ces méthodes : la méthode des Machines à Vecteurs Supports (SVM), les Réseaux de Neurones. Ces méthodes ont pour but d'identifier les classes auxquelles appartiennent les zones d'intérêts à partir de certains descriptifs.

2.5.1 La classification supervisée

La classification supervisée est une technique de l'apprentissage supervisé, qui a pour objectif de prédire une classe correcte sur des données futures. Comme son nom l'indique, la classification supervisée nécessite une supervision de la part d'un opérateur (ou d'un analyste) dans laquelle, « il identifie des échantillons assez homogènes de l'image qui sont représentatifs de différents types de classes d'information. » [33]. Les meilleures discriminations entre les classes est un problème majeur de la classification supervisée. La classification supervisée ou bien « un apprentissage constitué de données libellées si les différentes classes sont connues a priori sert à regrouper les individus de chaque classe dans deux échantillons : des échantillons d'apprentissage pour faire la dernière phase du système SAD qui est la classification. Les autres échantillons, de test servent à valider la classification en évaluant sa pertinence à travers le taux d'individus bien classés [3]. A titre d'exemples, nous citons : Les méthodes les plus réputées sont l'analyse discriminante linéaire, la régression logistique, les réseaux de neurones.

2.5.2 La classification non supervisée

La classification non supervisée ou bien le (Clustering) est utilisée dans le cas de données non libellées. L'objectif principal de cette méthode est de rassembler l'hétérogénéité des données dans des espaces différentes les plus homogènes possible entre eux et la bonne séparation entre ces espaces en fonction d'un critère de similarité (règle d'auto-organisation) défini a priori. La classification non supervisée, est une des tâches de l'apprentissage non supervisé. L'algorithme de Clustering utilise comme entrée un ensemble d'apprentissages de Données complètement non libellées, dont le but est de séparer les données en classes, sans aucune supervision [32]. Il existe de nombreux algorithmes de Clustering, citons : Kmeans, les k plus proches voisins, la classification hiérarchique ascendante, etc. [34].

2.5.3 La classification semi-supervisée

L'apprentissage semi-supervisé trouve ses racines dans les problèmes d'apprentissage en présence de données manquantes. Ainsi de nombreux travaux ont été effectués à ce sujet à la fin des années 1970 [1]. L'apprentissage semi-supervisé est une classe de techniques d'apprentissage automatique qui utilisent un ensemble de données étiquetées et non étiquetées. Il se situe ainsi entre l'apprentissage supervisé qui n'utilise que des données étiquetées et l'apprentissage non supervisé qui n'utilise que des données non étiquetées. Il a été démontré que l'utilisation de données non étiquetées, en combinaison avec des données étiquetées, permet d'améliorer significativement la qualité de l'apprentissage. Un autre intérêt provient du fait que l'étiquetage de données nécessite l'intervention d'un utilisateur humain. Lorsque les jeux de données deviennent très grands, cette opération peut s'avérer fastidieuse [35].

2.6 Conclusion

Dans ce chapitre, nous avons vu l'état de l'art concernant le traitement des images mammaire et quelques vues générales sur les descripteurs de bas niveau comme la texture, la couleur, et la forme, nous avons conclu par des notions sur les différentes méthodes de classification dans le domaine du traitement d'images médicales. Nous présentons dans le chapitre suivant les principes de la caractérisation.

Chapitre 3

Caractérisation Des Masses Mammaires

3.1 Introduction

La caractérisation est une tâche difficile qui a pour objectif de faire ressortir l'information contenue dans les images mamographiques en terme de forme et de texture, ces derniers sont définis comme une connaissance des paramètres importants sur lesquels le radiologue se base pour analyser un mammogramme et différencier les tissus normaux des tissus pathologiques .La forme est l'un des attributs de bas niveau les plus utilisés pour décrire la structure géométrique du contenu visuel. Même si la caractérisation du contenu géométrique s'est avérée complexe, plusieurs primitives géométriques ont montrées leurs intérêts dans les systèmes de recherche.

Dans ce chapitre, nous allons commencer par l'état de l'art sur les différents paramètres utilisés, puis nous détaillons comment les informations visuelles de texture et de forme, sont extraites à partir des images et de quelle manière les outils d'intelligence les utilisent pour rechercher l'information.

3.2 L'état de l'art

La texture est une composante riche en information, Ou bien un paramètre utilisé pour la compréhension et l'interprétation d'image. L'importance de l'information texture dans l'interprétation d'images montre l'intérêt que l'on lui porte dans l'analyse d'images. On se base dans ce cas, sur les méthodes statistiques du second ordre : le calcul de la matrice de cooccurrence.

- Haralick en 1973 [31] propose un outil statistique connu sous le nom matrice de co-

occurrence, elle permet de calculer la distribution des niveaux de gris dans l'image tout en préservant les interactions spatiales entre les pixels. Sharma a montré que les résultats obtenus par l'analyse des matrices de cooccurrences sont bons en termes de discrimination de textures en niveaux de gris [31]

- En ce qui concerne les micro-calcifications [8] ont confirmé que l'utilisation des caractéristiques de texture comme l'analyse des régions avoisinantes, la dépendance spatiale des niveaux de gris, les statistiques de longueurs sur les niveaux de gris ou encore la différence de niveaux de gris [8] sont les mieux adaptés.
- Dans le cas des opacités, la caractérisation repose sur la forme de l'objet [8] A titre d'exemples, On peut citer la compacité, les moments centrés, le moment de Hu ou encore le périmètre. [36,37] ont proposé des études sur l'analyse des distributions de la longueur radiale normalisée (distance entre le centroïde de la lésion et les points du contour) et des longueurs de cordes normalisées. La performance de ces mesures ne permet que de caractériser les résultats de la segmentation. Pour cette raison, la qualité de cette dernière influe sur la phase d'extraction des caractéristiques.[38] ont utilisé des mesures sur l'intensité par exemple l'analyse de la distribution des niveaux de gris dans la lésion supposée, sur son contour ou dans son voisinage. Par contre, L'utilisation des mesures qui sont invariantes par changement de contraste [39] a apporté des informations complémentaires. L'indice de spiculation est un caractère important pour indiquer que cette lésion est une lésion maligne. Par conséquence, une séquence d'études permettant d'évaluer à quel point la lésion est spiculée ont été proposées [40].Particulièrement, elles reposent sur l'analyse des orientations locales des contours [41]. [42] ont proposé la décomposition en ondelettes,[43]a utilisé l'analyse fractale pour évaluer la complexité d'un contour.

3.3 L'extraction des caractéristiques

ils existent deux types d'extraction (voir figure 3.1) :

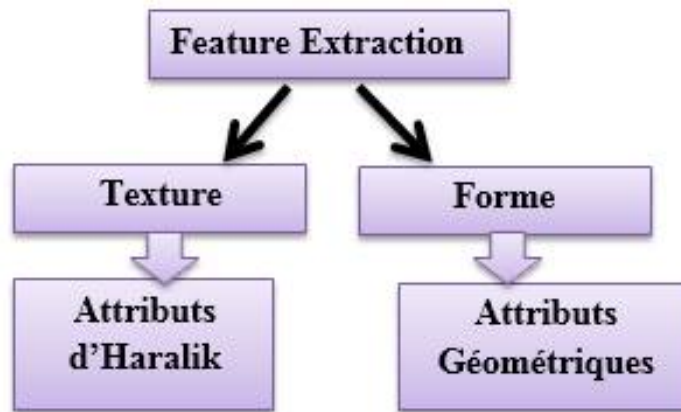


FIGURE 3.1 – Extraction des caractéristiques.

Il y a plusieurs méthodes pour la caractérisation, nous choisissons parmi ces méthodes les attributs d'Haralik de texture et les attributs géométriques de forme.

3.3.1 L'extraction d'attributs texturaux dans les images en niveau de gris

La littérature est très riche à ce sujet, on peut citer les trois familles principales :

- les approches structurales.
- les approches statistiques.
- les approches fréquentielles.

Puisque les images mammographiques possèdent des textures aléatoires et hétérogène, nous avons utilisées les approches statistiques.

3.3.1.1 Statistiques du premier ordre

L'analyse de texture par la méthode statistique du premier ordre permet de mesurer la probabilité empirique $p(n)$ du niveau de gris n ou l'histogramme des intensités noté *Histo*. Donner par la formule suivante :

$$Histo(i) = \frac{1}{NG} \sum_{p=1}^{NG} I(p) \equiv i \quad (3.1)$$

Avec I : l'intensité de l'image, NG : nombre de pixels et $i \in 1, 2, \dots, NG$ on cite les trois principaux statistiques du premier ordre :

— **La moyenne :**

$$Moyenne = \frac{1}{NG} \sum_{i=1}^{NG} i.Hist(i) \quad (3.2)$$

— **L'écart-type :**

$$\sigma_{Histo} = \sqrt{\frac{1}{NG} \sum_{i=1}^{NG} (i - Moyenne)^2 \cdot Histo(i)} \quad (3.3)$$

— **La variance :**

$$Variance = \frac{1}{N} \sum_{p=1}^N (I(P) - Moyenne)^2 \quad (3.4)$$

— **Les moments d'ordre k :**

$$\mu_k = \sum_n (n^k) p(n) \quad (3.5)$$

Cette description ne s'avère plus efficace, par ce qu'elle est basé sur l'histogramme donc n'exploite que l'information niveau de gris. L'inconvénient majeur c'est que cette méthode ne prend pas en compte les relations de voisinage entre pixels.

3.3.1.2 Statistiques du deuxième ordre

Les indices d'Haralick :

Dans le domaine de traitement d'image, la matrice de cooccurrence à niveau de gris GLCM est l'une des méthodes statistiques du deuxième ordre qui apporte une masse très importante d'informations, elle n'est pas directement exploitée pour caractériser les textures. Les utilisateurs préfèrent donc extraire de ces matrices des attributs afin de réduire la quantité d'informations à manipuler, tout en conservant la pertinence de ces descripteurs [31]. C'est pour cette raison que nous avons choisi d'utiliser les 13 premiers attributs d'Haralick extraits à partir de ces matrices [31]. Cette matrice Introduite initialement par Haralick en 1973, elle consiste à identifier le nombre de fois qu'un couples de pixels présentent dans une image en niveau de gris séparés par une distance d selon quatre directions $\theta = (0^\circ, 45^\circ, 90^\circ, 135^\circ)$. C'est une matrice carré de taille $N \times N$ (si N est la valeur maximale des niveaux de gris). Pour chaque (i, j) de Pd on construit une matrice, tel que Pd est le nombre d'occurrences des paires de niveaux de gris selon une certaine distance .Formellement :

$$pd(i, j) = ((r, s), (t, v)) : I(r, s) = i, I(t, v) = j \quad (3.6)$$

L'intérêt d'utiliser les approches statistiques comme la matrice de cooccurrence est l'aide pour fournir des informations précises sur la localisation relative des pixels voisins dans une région de l'image. Les directions d'analyse dans GLCM sont généralement (voir figure 3.2) :

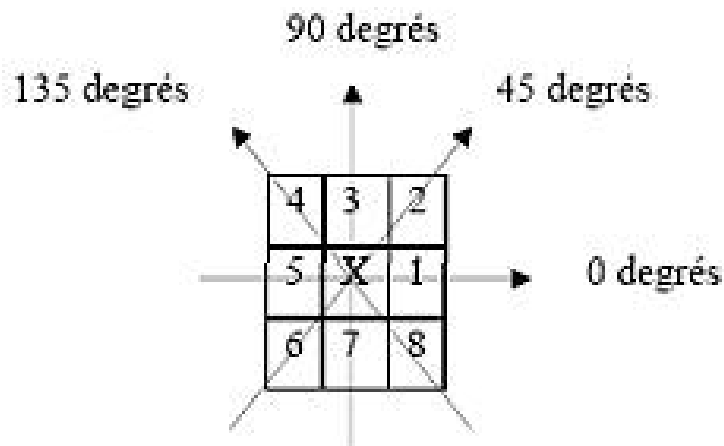


FIGURE 3.2 – Plus proches voisins du pixel 'X' selon les quatre directions.

	0	1	2	3
0	0	3	4	2
1	2	1	3	4
2	0	3	2	1
3	2	1	0	3

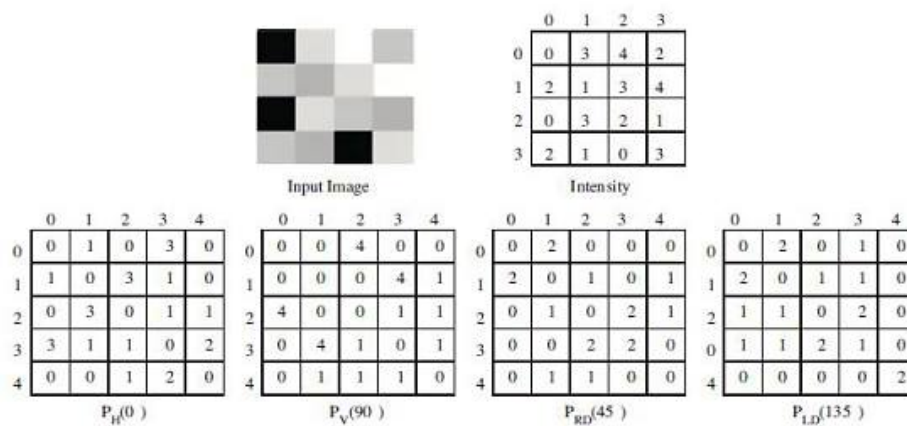


FIGURE 3.3 – la matrice de cooccurrence avec $D=1$ et $P_H(0^\circ), P_V(90^\circ), P_{RD}(45^\circ), P_{LD}(135^\circ)$.

3.3.1.3 Matrice de cooccurrence :

Algorithme :

- Construit les matrices de cooccurrence pour des directions données et des distances données.
- Calculer les vecteurs de caractéristique de la texture(nous avons gardé les 21 attributs de texture).

3.4 Attributs d'Haralik

Haralik dans son article "Textural features for image classification", a défini quatorze attributs de texture extraits à partir des matrices de cooccurrence. Ces attributs sont les suivants [31] :

1. Second moment angulaire (Energie) :

$$f1 = \sum_i \sum_j (P_{ij}(\sigma, \theta)^2) \quad (3.7)$$

Le second moment angulaire (ou énergie) permet de calculer l'uniformité de la texture donc mesure l'homogénéité de contenu interne de l'image : les fortes valeurs signifie que la distribution des niveaux de gris est constante. Ces fortes valeurs d'énergies sont obtenues pour les matrices de cooccurrence $P(\sigma, \theta)$ lorsque (σ, θ) correspond à la période. Par contre, les faibles valeurs indiquent des images moins uniformes.

2. Contraste :

$$f2 = \sum_i \sum_j ((i - j)^2 P_{ij}(\sigma, \theta)) \quad (3.8)$$

Le contraste (ou inertie) calcule les variations locales. Un fort contraste possède des valeurs plus élevé si ces variations sont importantes. Ce paramètre est fortement non corrélé à l'énergie.

3. Corrélation :

$$f3 = \frac{\sum_i^{Ng} \sum_j^{Ng} (ij) P_{ij}(\sigma, \theta) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.9)$$

Où μ_x, μ_y, σ_x et σ_y sont respectivement les moyennes et les écarts type de P_x et P_y . Ce paramètre indique la présence ou l'absence de certaines colonnes de la matrice qui sont égales, c'est-à-dire s'il y a des dépendances linéaires dans l'image. En effet, plus la corrélation est importante plus les valeurs sont uniformément distribuées dans la matrice de cooccurrences.

4. Variance :

$$f4 = \sum_i^{Ng} \sum_j^{Ng} (i - \mu)^2 P_{ij}(\sigma, \theta) \quad (3.10)$$

Où μ est la moyenne de tous les coefficients de la matrice ; La variance mesure l'hétérogénéité (la répartition des niveaux de gris autour de la valeur moyenne) de la texture. Elle augmente lorsque les niveaux de gris différent de leur moyenne. La variance est indépendante du contraste.

5. **Moment différentiel inverse :**

$$f5 = \sum_i^{Ng} \sum_j^{Ng} \frac{P_{ij}(\sigma, \theta)}{1 + (i - j)^2} \quad (3.11)$$

Ce paramètre a un comportement inverse de celui du contraste. En effet, plus la texture d'une région est homogène plus le moment différentiel inverse est élevé.

6. **Moyenne des sommes :**

$$f6 = \sum_i^{2Ng} iP_{x+y}(i) \quad (3.12)$$

7. **Variance des sommes :**

$$f7 = \sum_i^{2Ng} (i - f8)^2 P_{x+y}(i) \quad (3.13)$$

8. **Entropie des sommes :**

$$f8 = - \sum_i^{2Ng} P_{x+y}(i) \log(P_{x+y}(i)) \quad (3.14)$$

9. **Entropie :**

$$f9 = - \sum_i^{Ng} \sum_j^{Ng} P_{ij}(\sigma, \theta) \log(P_{ij}(\sigma, \theta)) \quad (3.15)$$

L'entropie permet de déterminer le degré de la complexité de l'image. Lorsque les valeurs de la matrice de cooccurrences sont toutes égales, l'entropie aura des valeurs élevées. Elle permet ainsi de caractériser la granulation de l'image. Par conséquent, plus la valeur de l'entropie est élevée et plus la granulation est grossière. L'entropie apporte des faibles valeurs si on a souvent les mêmes couples de pixels. C'est un indicateur de désordre. L'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire (sans structure apparente). Elle est fortement corrélée (par l'inverse) à l'énergie

10. **Variance des différences :**

$$f10 = \text{variance}(P_{x-y}) \quad (3.16)$$

11. **Entropie des différences :**

$$f11 = - \sum_i^{Ng-1} P_{x-y}(i) \log(P_{x-y}(i)) \quad (3.17)$$

12. Information sur la corrélation :

$$f_{12} = \frac{H_{XY} - H_{XY1}}{\max(H_X, H_Y)} \quad (3.18)$$

Où H_X et H_Y sont respectivement les entropies de P_x et P_y .

$$H_{XY} = - \sum_i \sum_j P(i, j) \log(P(i, j)) \quad H_{XY1} = - \sum_i \sum_j P(i, j) \log(P_x(i)P_y(j))$$

13. Information sur la corrélation :

$$f_{13} = (1 - \exp[-2(H_{XY2} - H_{XY})])^{\frac{1}{2}} \quad (3.19)$$

Où :

$$H_{XY2} = - \sum_i \sum_j P_x(i)P_y(j) \log(P_x(i)P_y(j))$$

14. Coefficient de corrélation maximal :

$$f_{14} = (i^{re} \text{ plus grande valeur propre de } Q)^{\frac{1}{2}} \quad (3.20)$$

Où :

$$Q = \sum_k \frac{P(i, k)P(j, k)}{P_x(i)P_y(k)}$$

Notons qu'Haralick a proposé ces 14 attributs de texture sans vraiment donner de précision sur leur signification. Cependant, les attributs les plus utilisés dans la littérature, comme l'énergie, le contraste, la corrélation, la variance, l'homogénéité et l'entropie.

Cependant, il est difficile d'interpréter ce que représentent les autres attributs, même de manière intuitive. Une étude complète serait nécessaire pour décrire ces attributs et mesurer la corrélation entre ces derniers, comme le précise Haralick [44].

A la suite d'Haralick, plusieurs auteurs comme Anys ont proposé d'autres attributs de cooccurrences dont les plus exploités sont présentés ci-dessous [44].

1. Moment diagonal :

$$f_{15} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \left(\frac{1}{2} |i - j| P[I](\sigma, \theta, (i, j)) \right)^{\frac{1}{2}} \quad (3.21)$$

2. Moyenne :

$$f_{16} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} i \times P[I](\sigma, \theta, (i, j)) \quad (3.22)$$

3. Cluster shade :

$$f_{17} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i + j - 2f_{16})^3 \times P[I](\sigma, \theta, (i, j)) \quad (3.23)$$

4. Cluster prominence :

$$f_{18} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i + j - 2f_{16})^4 \times P[I](\sigma, \theta, (i, j)) \quad (3.24)$$

5. Différence inverse :

$$f_{19} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{P[I](\sigma, \theta, (i, j))}{1 + |i - j|} \quad (3.25)$$

6. Dissimilarité :

$$f_{20} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} |i - j| \times P[I](\sigma, \theta, (i, j)) \quad (3.26)$$

7. Covariance :

$$f_{21} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i - \mu_x)(j - \mu_y) \times P[I](\sigma, \theta, (i, j)) \quad (3.27)$$

8. Probabilité maximale :

$$f_{22} = \max_{0 \leq i, j \leq Ng-1} (P[I](\sigma, \theta, (i, j))) \quad (3.28)$$

9. Troisième moment angulaire :

$$f_{23} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (P[I](\sigma, \theta, (i, j)))^3 \quad (3.29)$$

La plupart de ces attributs ont une signification semblable à celle des attributs d'Haralick. Cependant, certains d'entre eux sont intéressants de part le fait qu'ils permettent de caractériser des aspects différents de la texture[44].

3.5 Les descripteurs de formes

Les descripteurs de forme sont utilisés pour décrire les propriétés locales aux zones composant l'image. Contrairement aux descripteurs de textures utilisés pour décrire globalement l'image, la forme est un descripteur de bas niveau et l'une des principales

caractéristiques qui décrivent le contenu d'une image, leur robustesse sera au fonction de l'exactitude de la préalable segmentation et elle est généralement mesurée par le degré d'invariance à la translation, rotation et au facteur d'échelle[45].

La description et la discrimination des formes est l'un des problèmes de la vision par ordinateur et de la reconnaissance des formes [46] Les principales qualités d'un descripteur est d'une part sa grande capacité de discrimination de sorte qu'il puisse grouper des formes similaires entre elles et d'autre part à séparer les formes dissimilaires dans des groupes différents [46] Des perturbations telles que le bruit, les distorsions et les occlusions doivent être prises en compte car ils peuvent changer la forme d'une manière complexe et subtile [46].

La représentation de la forme est une tâche difficile, c'est parce que quand un objet en 3-D du monde réel est projeté sur un plan d'image en 2-D, une dimension de l'information de l'objet est perdue. En conséquence, la forme extraite de l'image ne représente que partiellement l'objet projeté[45].

Les descripteurs de formes devisent en deux grands types principaux ; approche basée contours et approche basée régions[47].

3.5.1 Approche basée contour

Prise en compte du contour seul. L'intérieur de la forme n'est pas considéré[47].

3.5.2 Approche basée région

Prise en compte des détails internes (trous ou nœuds)(voir figure 3.4). Les caractéristiques morphologiques décrivent la forme d'une structure qui représente des méthodes de préservation de l'information pour reproduire exactement la forme à partir de certains attributs. Ce qui justifie notre choix que ces attributs sont les plus significatifs pour caractérise une masse mammaire.

Généralement, ces attributs sont caractérisés par deux grands types : les caractéristiques spécifiques et les caractéristiques géométriques. Nous présentons dans ce qui suit quelques méthodes de description de la forme Géométriques.

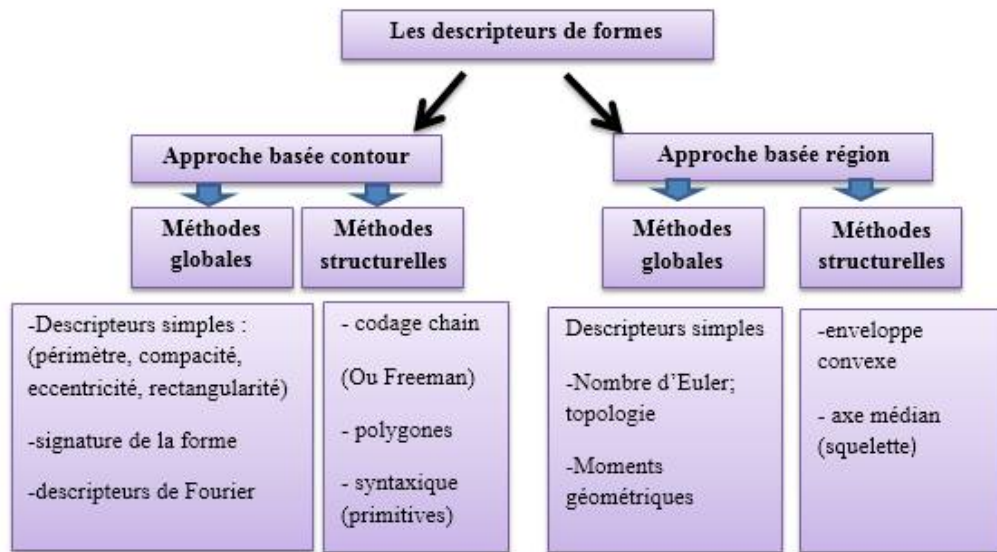


FIGURE 3.4 – Classification des techniques de la représentation de forme[47].

3.5.3 Attributs géométriques

Les attributs géométriques classiques se calculent sur des régions R_i homogènes. Nous citons par la suite :

1. L'aire :

La surface (aire) peut être obtenir par un simple comptage des pixels contenus dans une lésion noté A :

$$A = \sum \sum b(i, j) \quad (3.30)$$

2. Le périmètre :

Le périmètre sera calculé par la somme de chaque contour de chaque masse noté P est aussi couramment utilisé dans le domaine de diagnostic du cancer du sein.

$$P = \sum \sum c(i, j) \quad (3.31)$$

3. Le centre de gravité :

Le centre de gravité est calculé comme étant la moyenne des coordonnées en x et y , la forme est représentée par toute la région de l'objet :

$$C_i = \frac{X_i}{n}, C_j = \frac{Y_i}{n} \quad (3.32)$$

Avec n : nombre de pixel (surface).et (x_i, y_i) sont les points contenus dans cette région.

4. La circularité :

La mesure de circularité peut alors se définir comme suit [3] : plus la masse est circulaire et régulière et proche de 1, plus elle est bénigne (voir figure 3.5).

$$C = \frac{\text{aire}(R \cap C_{EQ})}{\text{aire}(R)} \quad (3.33)$$

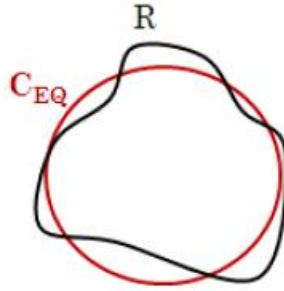


FIGURE 3.5 – Calcul de la mesure de circularité.

5. La rectangularité :

Un autre descripteur de forme intitulé rectangularité ou boîte englobante est souvent utilisé pour décrire le taux de rectangulaire et d'allongement d'une région [3].

$$Rec = \frac{\text{aire}(R)}{\text{aire}(B_E)} \quad (3.34)$$

On désigne par R la région étudiée et BE la boîte englobante qui est définie par le plus petit rectangle contenant la lésion étudiée. La mesure de rectangularité notée Rec peut alors se définir comme suit (voir figure 3.6) : En effet, la valeur de

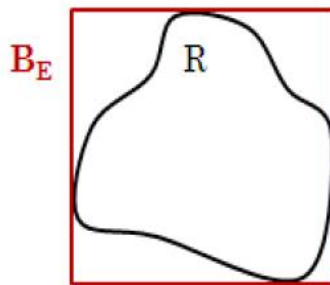


FIGURE 3.6 – Calcul de la mesure de rectangularité.

rectangularité d'un rectangle parfait et droit est égale à 1 alors que la rectangularité de ce même rectangle mais considéré incliné déférente de 1. Ainsi, ce descripteur est sensible à la rotation [3].

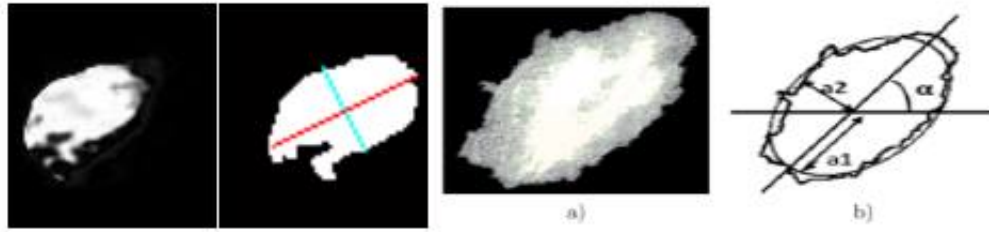


FIGURE 3.7 – l'ellipse équivalente d'un objet.(a) l'image originale ,(b) le résultat.

6. La rectangularité modifiée :

La formule de la rectangularité modifiée $MRect$ reste toujours le rapport entre l'aire de l'objet et l'aire de sa boîte englobante. Dans le contexte de la nouvelle formulation de la rectangularité ($MRect$) et quelque soit l'orientation de l'objet, plus la forme de l'objet est rectangulaire, plus l'entité $MRect$ tend vers 1. La rectangularité n'est pas invariante par rotation .Afin de palier à cette sensibilité à la rotation, on considère la boîte englobante minimale selon la direction de l'objet [3](figure 3.7).

7. l'asymétrie :

$$rapport = \frac{(A \cap B)}{(A \cup B)} \quad (3.35)$$

Avec : A : est la partie haut/gauche, B : la partie bas/droite. Plus ce rapport est proche de 0 plus la masse est bénigne.

La procédure pour calculer le rapport d'asymétrie ($R2$) :

- Création du masque binaire A qui représente la masse mammaire.
- Calcul le centre de gravité de la masse et les axes de symétrie.
- Devisions de la masse en 02 parties (droite et gauche).
- Rotation de 180% d'une partie de bw par rapport aux 02 axes de symétrie pour obtenir le masque binaire B .
- Calcul du rapport d'asymétrie ($R2$)pour les deux axes :

$$(R2) = \frac{(A \cap B)}{(A \cup B)}$$

8. La compacité :

La compacité invariant aux transformations géométriques telles que la translation, la rotation, il y a plusieurs formule qui indique la compacité .Elle est notée par $Comp$ et donnée par La formule :

$$Comp = \frac{P^2}{A} \quad (3.36)$$

Ou P est le périmètre de l'objet et A est la surface en pixels. Si la compacité égale à 1 donc l'objet est un cercle et attribue à la forme irrégulière une valeur plus élevée.

Ou : A est la surface en pixel.

9. **Convexe (Convex hull) :**

Une région R est convexe si et seulement si pour deux points quelconques $x_1; x_2 \in R$, la ligne entière du segment $x_1 x_2$ est à l'intérieur de la région. [45] Convex hull d'une région est la région convexe H la plus petite qui satisfait la condition

$$R \subset H$$

.La différence $H - R$ est appelé « carence convexe » D d'une région R. Une meilleure représentation de la forme peut être obtenue par un procédé récursif, ce qui conduit à un arbre de concavités[45](figure 3.8).

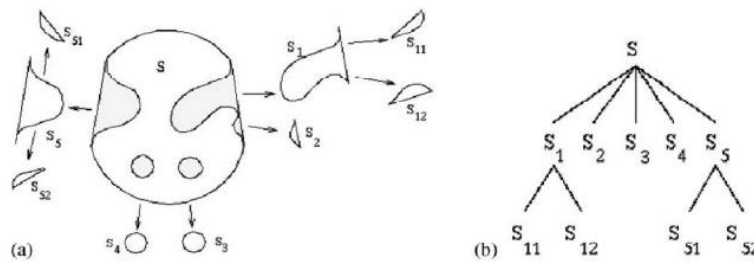


FIGURE 3.8 – (a) Convexe et ses concavités ; (b) l'arbre de concavité [45].

$$Comp1 = \frac{convex\ hull}{A} \tag{3.37}$$

$$Comp2 = \frac{maxRadius}{minRadius} \tag{3.38}$$

D'où : maxRadius est la largeur de l'axe majeur de la zone d'intérêt. minRadis est la largeur de l'axe mineur de la zone d'intérêt.

$$Comp3 = \frac{4 \times \Pi \times A}{P^2} \tag{3.39}$$

D'où : A est la surface en pixel. P est le périmètre.

10. **L'excentricité :**

$$E_x = \frac{la\ longueur\ d'axe\ majeur}{l'axe\ mineur} \tag{3.40}$$

11. **Nombre d'Euler :**

représente les différences entre nombre de composantes connexes et nombre de trous (Très utile pour reconnaître les figures géométriques).

12. L'axe médian (Medial axis) :

Comme convexe, un squelette d'une région peut également être utilisé pour la description et la représentation d'une forme. Un squelette peut être défini comme un ensemble connexe de lignes médianes. L'idée de base du squelette est de supprimer les informations redondantes tout en conservant uniquement les informations concernant la topologie de la structure d'objet qui peut aider à la reconnaissance [45]. La ligne dans la figure est le squelette de la forme. Le squelette peut alors être décomposé en segments et représenté par un graphe selon certains critères (figure 3.9).

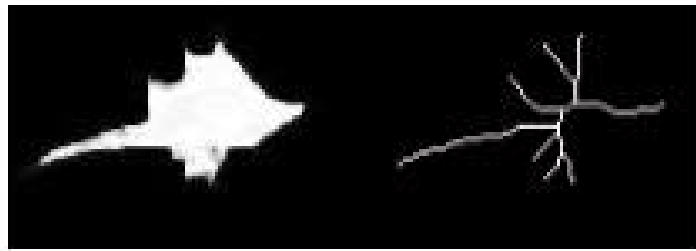


FIGURE 3.9 – Construction de l'axe médian d'une masse maligne.

Propriétés : Le squelette doit conserver les relations de connexité (voir figure 3.10) :

- même nombre de trous par composante connexe.
- même nombre de trous par composante connexe.



FIGURE 3.10 – Exemple montrant les propriétés du squelette.

13. **Invariance aux transformations affines** Translation, rotation et homothétie.
14. **Continuité :** Une petite modification de la forme originale doit induire une petite modification du squelette(voir figure 3.11)

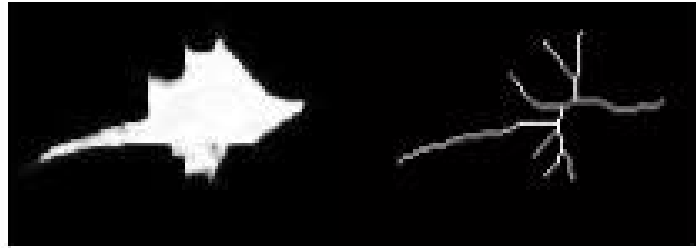


FIGURE 3.11 – Exemple montrons la continuité du squelette.

15. **Continuité** : elle est très sensible à de faibles variations de contour.
16. **Réversibilité** : Le squelette doit permettre de retrouver la forme originale.

3.6 Conclusion

Nous avons présenté dans ce chapitre, les différents descripteurs de textures et des formes dédiés pour la caractérisation des masses mammaires dans la littérature, assurent l'invariance par translation, par rotation et par changement d'échelle. Dans le chapitre suivant, nous procédons à l'évaluation expérimentale de ces résultats afin de vérifier leur performance dans le cadre de la classification des masses mammaires dans une base d'images mammographiques bien déterminée.

Chapitre 4

Résultats et Discussion

4.1 Introduction

La classification est une problématique qui est traitée dans plusieurs communautés de recherche .Elle découvre et enrichisse mutuellement : les statistiques, les reconnaissances de formes (RDF), l'apprentissage automatique. L'objectif de l'apprentissage artificiel est de réaliser un modèle appelé classifieur, ce dernier permet de détecter la similarité entre les caractéristiques de la même classe et les dissemblances entre les caractéristiques de classes différentes, à titre d'exemple, on peut classifier les masses mammaires en deux classes : malignes ou bénignes.

Dans ce chapitre, nous présentons le cadre général de l'apprentissage pour développer les différents classifieurs à savoir le réseau de neurones multicouches (RNMC) et les séparateurs à vaste marge (SVM). La phase de caractérisation correspond à l'extraction d'attributs en termes de texture et de forme contenus à l'intérieur de l'image. Nous nous proposons trois (03) scénarios de caractérisation :

- vecteur de caractéristique avec texture seul.
- vecteur de caractéristique avec forme seul.
- vecteur de caractéristique avec textures et forme.

Ensuite une étude comparative a été effectuée pour montrer l'intérêt de chaque type de caractérisation.

4.2 Classification neuronal :

Un classifieur neuronal multicouche reçoit des données par l'intermédiaire d'une couche réceptrice « de neurones », traite ces données avec ou sans l'aide d'une ou plusieurs couches « cachées » contenant un ou plusieurs neurones et produit une valeur

à la sortie [48].

Une couche d'entrée afin d'intégrer des valeurs descriptifs du problème traité, le nombre de neurones dépend du nombre de valeurs, une couche ou plusieurs couches cachées qui nous permet d'effectuer des traitements sur les données d'entrées en utilisant des fonctions d'activation, ce traitement a pour but d'extraire des données qui seront par la suite utilisées dans la couche de sortie. Pour ne pas augmenter la complexité du réseau, nous décidons d'utiliser une seule couche cachée. Une couche de sortie qui sert à identifier le résultat de la classification, dans notre cas nous avons un seul neurone de sortie (correspondant à une masse bénigne ou maligne). (voir figure 4.1).

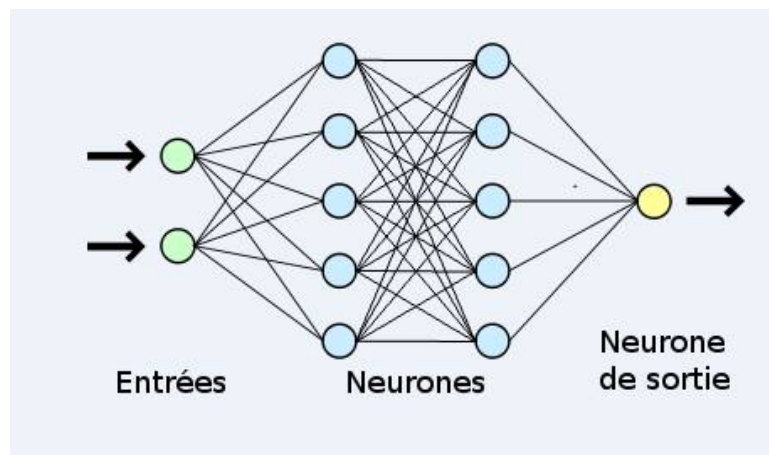


FIGURE 4.1 – perceptron multicouches.

4.3 Classification par SVM

Ce sont des algorithmes d'apprentissage initialement construits pour la classification binaire. L'idée est de rechercher une règle de décision basée sur une séparation par un hyperplan de marge optimale. Méthode relativement récente qui découle de premiers travaux théoriques de Vapnik et Chervonenkis en 1995 [58].

4.3.1 Principe de l'algorithme :

Pour deux classes d'exemples donnés, le but de SVM est de trouver un hyperplan qui peut séparer les données et maximiser la distance par rapport à l'hyperplan. Les points les plus proches sont seuls utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support (voir figure 4.2). Il est évident qu'il existe une multitude d'hyperplans valide, mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. Nous allons donc en plus chercher parmi les hyperplans valides,

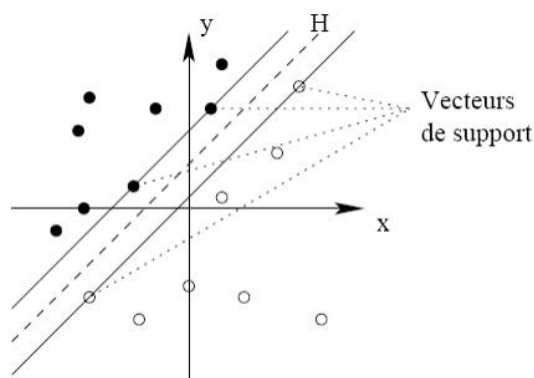


FIGURE 4.2 – représente un hyperplan.

celui qui passe « au milieu » des points des deux classes d'exemples. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr ». En effet, une distance maximale de marge, c'est-à-dire la distance entre l'hyperplan et les données d'apprentissage décrit un bon séparateur.

4.4 Présentation de la base DDSM d'images mammographiques :

4.4.1 Le choix de la base DDSM :

Nous avons choisi la base de données *DDSM* reconnue en anglais Digital Database for Screening Mammography (DDSM) [3]. C'est une base de données numérique construite à partir de films numérisés. Elle a été rassemblée par un groupe de chercheurs de l'université Sud de Floride et a été largement utilisée par la communauté scientifique dans le domaine du cancer du sein, cette dernière contient 2620 cas recueillis auprès de l'hôpital Massachusetts Général Hôpital (*MGH*).

Les masses mammaires apparaissent sous plusieurs formes spiculées ou circonscrites, les formes des masses bénignes sont généralement : ronde, lobulaire ou ovale. Par contre, les masses malignes sont de formes irrégulières ayant des frontières non définies est suspectes. Donc, la base DDSM emploie le même lexique standardisé par l'American College of Radiology dans le BIRADS (voir figure 4.3). La base DDSM doit comprendre tous les cas de figure possibles et connus par les chercheurs. Cette dernière a une capacité de recherche conçue pour nous permettre d'identifier les cas qui répondent aux critères spécifiés tels que la normale / cancer / bénigne. Chaque case est composée de quatre clichés (l'incidence oblique externe (MLO) et l'incidence Cranio Caudale (CC) de chaque sein).

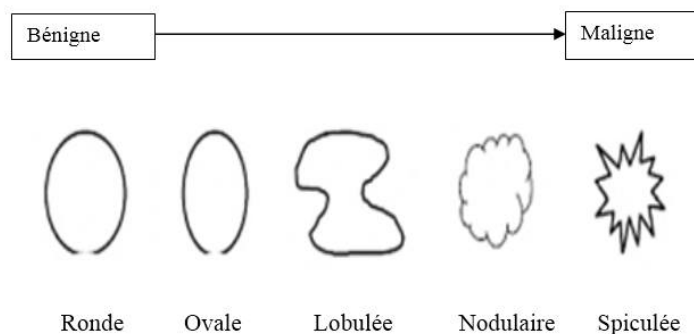


FIGURE 4.3 – Représentation de l'évolution du signe de malignités des masses mammographiques.

La figure (voir figure 4.4) montre la case 0024 située dans le dossier cancer 02, avec les annotations du médecin des deux seins droit (RMLO, RCC) et gauche (LMLO, LCC).

La sous base considérée dans notre travail est constituée de 90 masses dont 52 bénignes et 38 malignes, sont sélectionnés afin d'être utilisées dans la partie expérimentale. Cette base est partitionnée en 60 images d'apprentissage et 30 images de test. Le tableau 4.9 récapitule les différents détails concernant la base utilisée (B désigne les masses bénignes et M les masses malignes) (tableau 4.1) .

Nombre total de masse	90(52B/38M)
Nombre de masses d'apprentissages	60
Nombre de masses de tests	30

TABLE 4.1 – Répartition de la base utilisée.

4.5 Le choix des descripteurs :

Nous avons utilisé les approches statistiques qui sont les mieux adaptées, par ce que les images mammographiques possèdent des textures aléatoires et inhomogènes. Tandis que, les approches structurales sont des méthodes d'analyse adaptées aux textures macroscopiques et régulières, les approches fréquentielles sont utilisés pour analyser les textures régulières et périodiques.

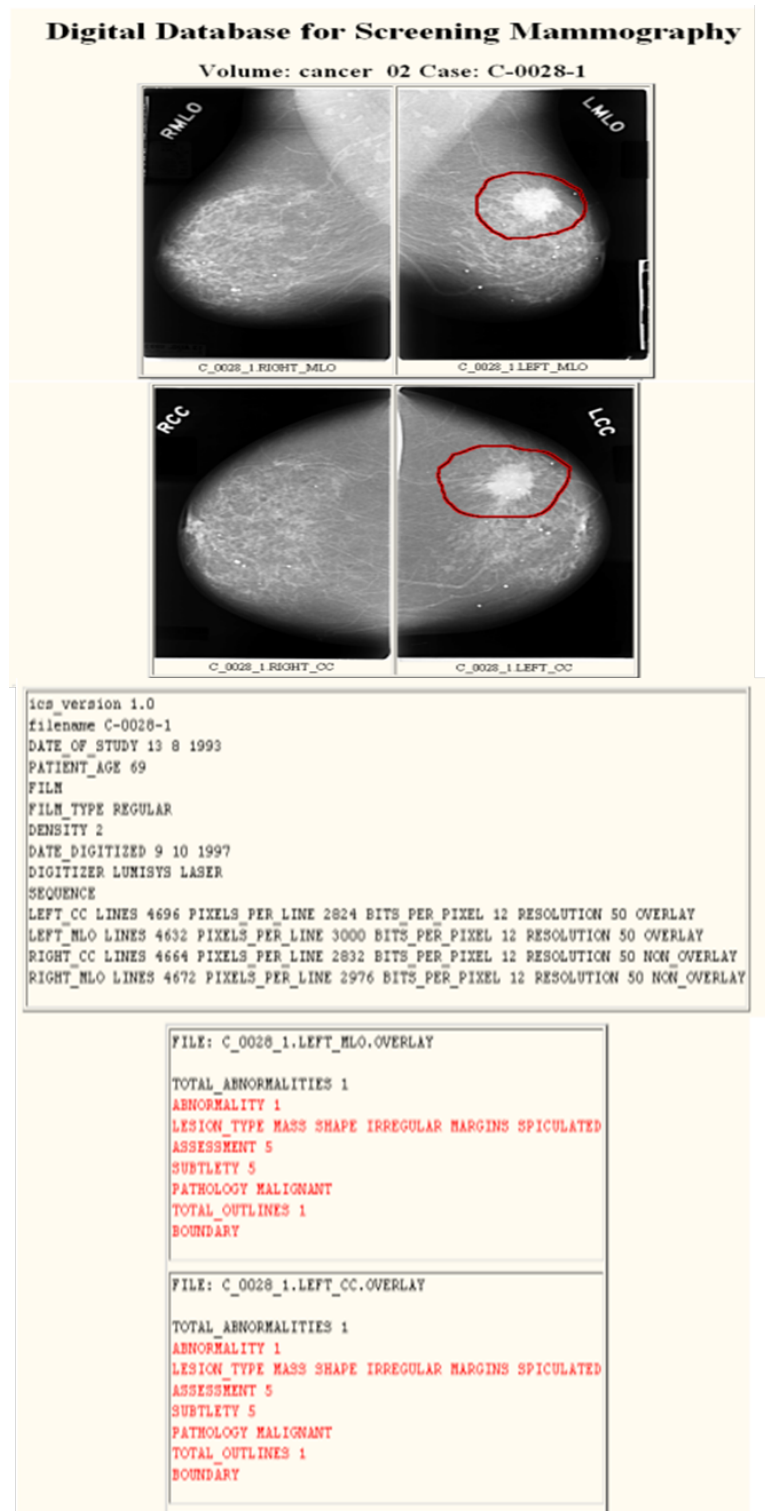


FIGURE 4.4 – Cas 0004 de la base DDSM avec annotations.

Les recherches antérieures ont montré que les descripteurs de forme sont plus appropriés pour l'analyse des masses mammaires .Leur degré de malignité augmente

essentiellement avec l'élévation de la complexité du contour [48]. La figure 3 montre les différentes formes de contour d'une masse selon un degré croissant de malignité.

4.5.1 La caractérisation de la masse mammaire :

En terme de malignité et de bénignité, l'analyse des images mammographiques permet de détecter, de localiser et de prévoir la gravité de la tumeur.

En exploitant les caractéristiques extraites de la masse, généralement regroupées en deux types : forme, texture (voir chapitre 2 et 3). L'utilité d'extraction de ces caractéristiques est de mieux guider les radiologues lors de leur diagnostic, afin d'établir le cadre de dépistage précoce du cancer du sein.

Nous avons utilisé 34 descripteurs de forme et de texture proposés dans la littérature à savoir 13 descripteurs géométriques :

- l'aire (A).
- le périmètre (P).
- la circularité (C).
- la rectangularité (Rect).
- la rectangularité modifiée (MRect).
- la compacité (Com).
- la convexité .
- Excentricité (Ex).
- L'étude des distorsions architecturales.
- le diamètre équivalent.
- la solidarité.
- EulerNumber.
- Les points terminaux du squelette.

21 descripteurs de textures (la matrice de cooccurrence) :

- Auto-corrélation.
- Contraste.
- Corrélacion.
- Corrélacion.
- Cluster Prominence.
- Cluster Shade.
- Energie.

- Dissimilarité.
- Entropie.
- Homogénéité.
- probabilité maximale.
- Moyenne des sommes.
- Variance des sommes.
- Variance des différences.
- Entropie des différences.
- Information sur la corrélation(1).
- Information sur la corrélation(2).
- Différence inverse normalisée.
- Le moment de différence inverse normalisée.
- Entropie des sommes.
- Moment diagonal/Variance des Sommes.

4.5.2 Présentation de notre modèle de traitement :

Après l'étape de segmentation qui est primordiale dans l'extraction de la masse, l'étape d'extraction de caractéristiques consiste à représenter l'image en deux familles qui sont : les attributs de texture et les attributs géométriques. En effet, la bonne détection de contour, améliore la description des lésions mammaires et par la suite le taux de la classification. Nous avons effectué plusieurs tests empiriques afin de garder ceux qui donnent le meilleur taux de classification(voir figure 4.5).

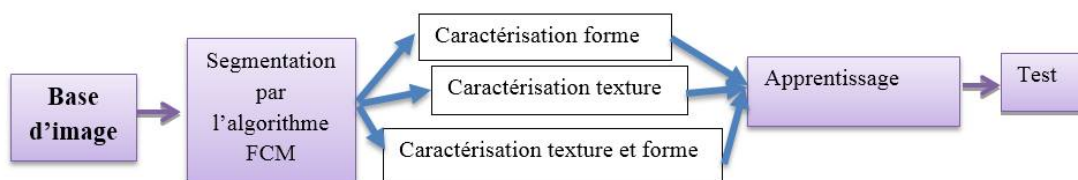


FIGURE 4.5 – chaîne de traitement abordée.

4.6 Les caractéristiques de la masse :

4.6.1 la forme :

Nous nous intéressons aux attributs géométriques basés sur des propriétés visibles à l'œil (la compacité, périmètre). C'est à dire l'image est considérée comme un espace géométrique.

Donc, après la binarisation de la masse, il est possible d'obtenir un certain nombre de mesures géométriques (voir figure 4.6) :

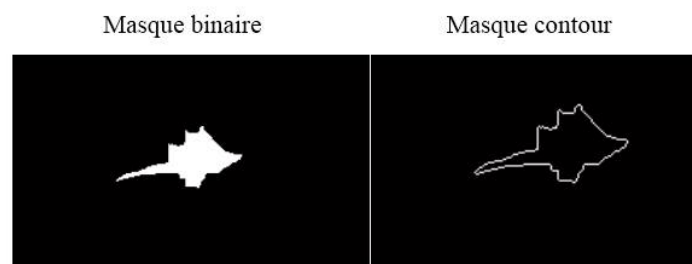


FIGURE 4.6 – Extraction de mesures morphologiques à partir du masque binaire ou de contour.

En mammographie la forme et la régularité des bords sont parmi les critères les plus utilisés pour la caractérisation des masses mammaires, La figure 4.7 montre la mesure de compacité, le périmètre, le Rapport d'asymétrie et le nombre du point terminaux du squelette. De deux types de masse (maligne et bénigne).

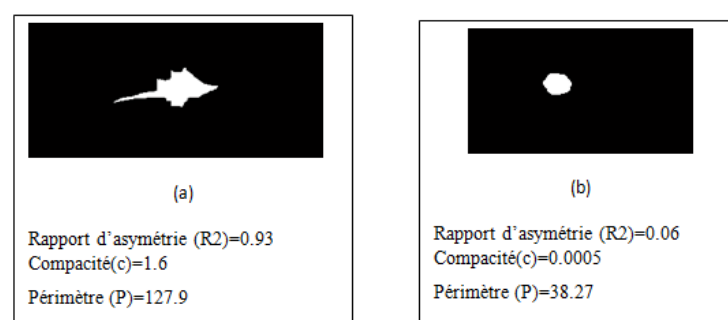


FIGURE 4.7 – Le rapport d'asymétrie, la compacité et le périmètre des masses mammaires. (a) masse maligne, (b) masse bénigne.

La discrimination est élevée, si on part de l'idée qu'un descripteur de forme doit être efficace. Et si on projette cette condition sur le descripteur de compacité, rapport

d'asymétrie, et l'axe médian, on peut dire que : D'après la compacité obtenue pour les deux masses (tableau 4.2) , la masse bénigne présente un aspect presque régulier ($C=0.0005$), tandis que la masse maligne a une valeur très loin de 1 (aspect irrégulier).ces résultats sont fortement liés aux résultats de la segmentation.

D'après le rapport d'asymétrie obtenu pour les masses (maligne/bénigne) (tableau 4.2) les masses bénignes tendent vers une forme ronde ($R2=0.06$), cet indice s'approche de 0 donc la lésion présente une certaine symétrie. Les masses malignes tendent vers une forme irrégulière ($R2=0.93$), cet indice de symétrie témoigne d'une asymétrie de la forme, cela justifie la probabilité de malignité de la masse.

Les points terminaux de l'axe médian (La figure 4.8) est un descripteur très prometteur et très fidèle au contour des masses mammaires. Le nombre de points terminaux est proportionnel à la complexité de la forme. D'après le tableau 4.2, ce descripteur est nécessaire pour pouvoir discriminer les masses malignes et des masses bénignes. Les points terminaux de l'axe médian satisfont la condition d'invariance aux transformations géométriques et changement d'échelle.

	Rapport d'asymétrie (R2)	Compacité(c)	Périmètre (P)	Nombre du point terminaux (NPT)
Masse maligne	0.93	1.6	127.9	8
Masse bénigne	0.06	0.0005	38.27	4

TABLE 4.2 – Récapitulatif de mesure d'attributs de forme


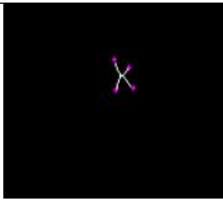

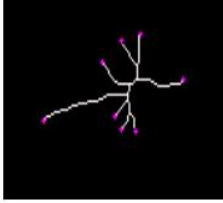
	Lésion	Squelette	Nombre de points terminaux
Bénigne			4
Maligne			8

FIGURE 4.8 – Les points terminaux du squelette.

Les descripteurs d'ordre général comme le périmètre, l'aire, la compacité, l'excentricité, la réctangularité et la réctangularité modifiée sont supposées croître en fonction de la complexité de la forme. Donc l'utilisation de ces derniers individuellement s'avère peu ou moyennement efficaces. En effet, l'aire et le périmètre d'une masse régulière de taille importante peuvent dépasser l'aire et le périmètre d'une masse irrégulière de petite taille.

Ces résultats s'expliquent par le fait que l'information apportée par ces descripteurs soit insuffisante pour décrire des formes complexes telles que les masses mammaires. A titre d'exemple, les mesures de l'aire A et du périmètre P .

Même si ces descripteurs géométriques sont insuffisants pour décrire la globalité des informations contenues dans une forme, ils demeurent toujours utiles pour apporter quelques détails supplémentaires concernant la forme en question. Plusieurs travaux de recherche ont montré que l'utilisation de ces descripteurs combinés avec d'autres plus adaptés permet généralement d'améliorer la caractérisation des formes [3].

4.6.2 la texture :

Une description comparative des masses bénignes et malignes a été réalisée : La valeur de l'énergie dans l'image maligne (0.97), est plus forte que celle dans l'image bénigne (0.02), Cette dernière signifie que la distribution des niveaux de gris est constante dans les images malignes, par contre, elle est moins uniforme dans les images bénignes (tableau 4.3).

Selon les résultats du contraste (tableau 4.3), les masses malignes possèdent des variations locales importantes par rapport aux images bénignes. L'indice de corrélation est utilisé pour déterminer si certaines colonnes de la matrice sont égales, c'est-à-dire s'il existe des dépendances linéaires dans l'image. Donc, les niveaux de gris des masses bénignes sont uniformément distribués dans la matrice de cooccurrences par rapport aux niveaux de gris des masses malignes (tableau 4.3).

La variance permet de savoir l'hétérogénéité de la texture. Les masses malignes portent des valeurs fortes par rapport aux masses bénignes ce qui justifie que la texture des masses malignes est hétérogène (tableau 4.3).

Plus le moment différentiel inverse est élevé plus la texture possède de régions homogènes cela signifie que les masses malignes ont des textures homogènes. Les masses bénignes possèdent des textures moins homogènes (tableau 4.3).

La complexité de l'image est mesurée par l'indice de l'entropie ou bien par l'indicateur de désordre. Si les valeurs de la matrice de cooccurrences sont égaux, l'entropie est

élevée ce qui est le cas des masses bénignes, par contre dans des masses malignes sont moins désordonnées (tableau 4.3).

Les attributs qui restent, permettent d'ajouter des informations supplémentaires concernant les degrés d'homogénéité, et de complexité de l'image, ainsi que la corrélation.

Attributs texture	Masse maligne	Masse bénigne
Auto-corrélation	1.47	1.11
Contraste	0.06	0.01
Corrélation	0,92	0.89
Corrélation	0,92	65.58
Cluster Prominence	260,23	4.8
Cluster Shade	19,31	0.003
Dis-similarité	0,012	0.99
Énergie	0,97	0.02
Entropie	0,07	0.99
Homogénéité	0,97	0.99
probabilité maximale	0,99	0.99
Moyenne des sommes	1,46	2.02
Sommes des variance	4.39	2.11
Sommes des entropie	5,63	0.01
Moment diagonal	0.98	1.09
Variance des différences	0,06	0.01
Entropie des différences	0,03	-0.71
Information sur la corrélation(1)	-0,71	0.16
Information sur corrélation(2)	0,28	0.99
Différence inverse normalisée	0,99	0.99
Le moment de différence inverse normalisée	0.99	0.99

TABLE 4.3 – Récapitulatif de mesure d'attributs de texture.

4.7 Classification :

L'objectif de cette application est de classer les masses comme étant des masses malignes ou bénignes. On fait appel à la classification supervisée. Elle consiste à affecter à chaque zone extraite contenue dans l'image une classe. Ce type de classification nécessite deux bases, une pour l'apprentissage et une autre pour le test, la répartition entre ces deux bases doit être faite expérimentalement.

4.7.1 Base d'apprentissage :

Elle représente 2/3 de notre base, la nécessité de la phase d'apprentissage sert à ajuster les paramètres du classifieur en minimisant l'écart entre la sortie du modèle et les mesures faites sur les exemples d'apprentissage. Cette tâche d'apprentissage nécessite une optimisation de la fonction de coût d'où la définition de cette dernière est primordiale, Afin de maximiser le taux de reconnaissance ou de minimiser l'erreur.

4.7.2 Phase de test :

Cette phase doit permettre l'affectation d'un nouvel objet à l'une des classes, au moyen d'une règle de décision intégrant les résultats de la phase d'apprentissage, afin de s'assurer la fiabilité du classifieur sur différents plans : Le taux de classification ou de reconnaissance, le taux d'erreurs, la sensibilité et la spécificité (la fiabilité des résultats dépend de la performance du classifieur).

Pour évaluer les performances de nos classifieurs (neural, SVM), nous déterminons le taux de sensibilité (Se), spécificité (Sp) et taux de classification (CC). Ces paramètres combinent entre les vrais positifs et les faux positifs. Ils sont définis comme suit [48] :

$$se = \frac{VP}{VP + FN}$$

$$sp = \frac{VN}{VN + FP}$$

$$CC = 100 \times \frac{VP + VN}{VP + VN + FP + FN}$$

- VP (Vrai Positif) : est le nombre de lésions malignes qui sont classées malignes.
- FP (Faux Positif) : est le nombre de lésions bénignes qui sont classées malignes.
- VN (Vrai Négatif) : est le nombre de lésions bénignes qui sont classées bénignes.
- FN (Faux Négatif) : est le nombre de lésions malignes qui sont classées bénignes.

4.8 Les résultats obtenus :

Dans cette étape nous avons utilisé une banque de données de 90 images (bénigne et maligne). La répartition de l'échantillon entre les deux phases se fait aléatoirement sur la banque de données pour chacune des 02 phases (2/3 pour la phase d'apprentissage et 1/3 pour la phase de test). Dans notre cas, nous cherchons à apprendre le modèle à classer les tumeurs (malignes ou bénignes) à partir des valeurs fournis à son entrée. Ces valeurs sont présentés sous forme de paramètre (éléments du vecteur d'entrée), La

sortie du modèle est la permittivité relative qui correspondre deux classes à savoir (1 : bénigne, 2 : maligne), nous avons obtenu les performances indiqués dans les Tableaux (tableau 4.4) :

L'évaluation des performances des deux classifieurs est estimée par le taux de classification, la sensibilité et la spécificité.

Dans la première expérimentation, nous avons utilisé les classifieurs SVM. Ces derniers ont donné des meilleures performances particulièrement dans la reconnaissance des cas malins où une sensibilité de 94.73% a été atteinte (tableau 4.4).

	Taux	sensibilité	spécificité
Forme	86.67%	81.25%	92.85%
Texture	90%	94.73%	81.81%
Texture et forme	86.67%	84.21%	90.90%

TABLE 4.4 – Classification par SVM.

Dans la deuxième expérimentation, nous avons utilisé le classifieur neuronal (tableau 4.5).

	Taux	sensibilité	spécificité
Forme	86.67%	100%	75%
Texture	64.28%	72.73%	68.82%
Texture et forme	80%	72.73%	84.21%

TABLE 4.5 – Classification neuronal.

La classification neuronale a donné des meilleures performances particulièrement dans la reconnaissance des cas malins où une sensibilité de 100% a été atteinte (tableau 4.5) en utilisant les attributs de forme.

D'après les résultats obtenus, nous remarquons que les méthodes de classification proposées donnent de bons résultats que ce soit pour la classification, neuronale ou SVM, Ceci est dû au fait que les parties segmentation et caractérisation ont été bien faites.

4.9 Conclusion :

Dans ce chapitre nous avons mis au point un modèle pour la segmentation, caractérisation et classification pour la caractérisation des masses mammaires .Pour cela, nous avons procédé après la phase de segmentation à une étape de caractérisation, en utilisant deux types de paramètres (texture et forme),les résultats obtenus de cette dernière ont été présenté à l'entrée de deux classifieur qui sont respectivement :SVM et RNMC. Dont lequel nous avons évaluée leurs performances à l'aide de trois critères à savoir : le taux de la classification, la sensibilité et la spécificité. Les résultats obtenus ont été prometteurs.

Conclusion générale et perspectives

Notre modèle de classification concerne l'étude des systèmes de reconnaissance automatique des pathologies en vue du diagnostic précoce du cancer du sein, notamment dans la phase de caractérisation des masses mammaires pour mieux guider les radiologues dans leur diagnostic, et par la suite nous avons achevé notre travail par la phase de classification pour évaluer la performance de notre modèle.

Les chercheurs dans le domaine de l'image se focalisent sur les étapes de segmentation et de caractérisation afin d'améliorer le résultat de classification des masses en malignes/bénignes. Un état de l'art détaillé concernant les parties de segmentation et caractérisation a été présenté.

Nous avons traité des masses mammaires extraites de la base de données DDSM classées en deux catégories : masses malignes et masses bénignes. Cette dernière est déjà traitée par des experts et la zone d'intérêt est entourée par un cercle rouge. La méthode de segmentation la plus fiable est d'utiliser des outils de traitement d'image comme exemple « Image J ».

La phase de caractérisation consiste à extraire les informations sur laquelle les décisions finales des radiologues se reposent à savoir les descripteurs géométriques de forme et les descripteurs de texture. Les recherches antérieures ont montré que ces descripteurs sont plus adaptés à la nature des lésions de type micro-calcifications alors que les descripteurs de forme sont plus appropriés pour l'analyse des masses mammaires.

La méthode de codage, proposée consiste à transformer l'information textuelle, exprimée en niveau de gris contenu dans les tissus d'un mammogramme. En effet, ces images codées sont plutôt riches en information par rapport aux images d'origine ce qui rend la détection et la caractérisation des masses plus performante et plus efficace. La meilleure performance est assurée par l'étude de l'influence de certains paramètres sur la caractérisation des pathologies mammaires. Nous avons évalué et testé les per-

formances de chaque classifieur en terme de sensibilité (Se), Spécificité (Sp) et le taux de classification correcte (CC).

Comme perspectives, nous proposons d'utiliser la combinaison de plusieurs classifieurs. Il s'agit précisément de remédier aux erreurs résultantes d'un seul classifieur. En effet, la coopération apporte toujours des améliorations par rapport à un seul classifieur. Aussi l'enrichissement de la phase de caractérisation des lésions mammaires pour le diagnostic précoce du cancer du sein (l'utilisation des moments géométriques : les moments centrés, les moments de Hu), pour améliorer les résultats.

Une classification semi-supervisé par S3VM et une segmentation par classification pixel-aire des masses mammaires peut être recommandée comme une autre partie de recherche.

Références bibliographiques

Reference :

- [1] Nawel et Nabiha, « *Une approche semi-supervisée pour la classification des images mammographiques* », publication, Tlemcen 2014.
- [2] Fatima EDDAOUDI, « *AIDE A LA DETECTION DES PATHOLOGIES DU SEIN PAR LES TECHNIQUES DE SEGMENTATION ET DE CODAGE DE TEXTURES DES IMAGES MAMMOGRAPHIQUES* », THÈSE DE DOCTORAT UNIVERSITÉ MOHAMMED V – AGDAL FACULTÉ DES SCIENCES ,Rabat 2012.
- [3] IMENE CHEIKHROUHOU, « *Description et classification des masses mammaires pour le diagnostic du cancer du sein* », thèse de l'UNIVERSITE D'EVRY VAL D'ESSONNE, Laboratoire d'Informatique, Biologie Intégrative et Systèmes Complexes,2012.
- [4] C Balleyguier et al, « *Comprendre et utiliser le CAD (Aide informatisée au diagnostic) en mammographie* », publication, paris 2005.
- [5] http://gyneco-obstetrique.hug-ge.ch/gynecologie/seno_echo_sein.pdf.
- [6] J.Brettes, C. Mathelin, B. Gairard, J. Bellocq. Cancer du sein. Paris : Elsevier Masson, 2007. 358 p. ISBN : 978-2-294-01813-8.
- [7] Mr hamza chekkaf, Mr ibrahim touil, « *segmentation des masses dans les images mammographique* », mémoire de fin d'études pour l'obtention du diplôme de master en informatique, 25 septembre 2012.
- [8] Giovanni Palma, « *Détection automatique des opacités en tomosynthèse numérique du sein* », Télécom ParisTech, 2010. French. <pastel-00005948>.*2
- [9] Zhou, X. H. et Gordon, R. (1989). Detection of early breast cancer : an overview and future prospects. *Critical Reviews in Biomedical Engineering (CRBE)*, 17(3):203–255.
- [10] G. Kom, A. Tiedeu, M. Kom, C. Nguemgne, J. Gonsu - Détection automatique des opacités dans les mammographies par la méthode de minimisation de la somme de l'inertie, *ITBM-RBM*, Vol. 26, Issues 5-6, Octobre-Décembre 2005, Pp. 347-356.
- [11] Heath, M. D. et Bowyer, K. W. (2000). Mass detection by relative image intensity. *International Workshop on Digital Mammography (IWDM)*. Medical Physics Publishing (Madison, WI).
- [12].Vachier, C. (1995). Extraction de caractéristiques, segmentation d'image et morphologie mathématique. Thèse de doctorat, Ecole nationale supérieure des mines de Paris.
- [13]. Bornefalk, H. (2005). Use of quadrature filters for detection of stellate lesions in mammograms. *Scandinavian Conference on Image Analysis (SCIA)*, pages 649–658.
- [14] S. Vijaya Kumar, M. Naveen Lazarus, C. Nagaraju - A Novel Method for the Detection of Microcalcifications Based on Multi-scale Morphological Gradient Watershed Segmentation Algorithm, *International Journal of Engineering Science and Technology*, Volume 2, Issue 7, Pages 2616-2622, 2010.
- [15] K. Thangavel, M. Karnan, R. Sivakumar, A. Kaja Mohideen - Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms, *International Journal on Graphics Vision and Image Processing*, 2006.
- [16] Rabi Narayan Panda, Bijay Ketan Panigrahi, Manas Ranjan Patro Feature Extraction for Classification of Microcalcifications and Mass Lesions in Mammograms, *International Journal of Computer Science and Network Security (IJCSNS)*, Volume 9, No 5, May 2009.

- [17]. P. sante. (2012) Passeport sante. <http://www.passeportsante.net/fr/>.
- [18]. L. F. Campos, A. C. Silva, and A. K. Barros, Diagnosis of Breast Cancer in Digital Mammograms Using Independent Component Analysis and Neural Networks. CIARP, 2005.
- [19]. N. Petrick, N. Sahiner, H. Chan, M. Helvie, S. Paquerault, and L. Hadjiiski, Breast cancer detection : Evaluation of a mass-detection algorithm for computer-aided diagnosis, Radiology, vol. 224, 2002.
- [20]. H. Hwang, B. Choi, H. Lee, H. Yoon, Y. Nam, and H. Choi, Multiresolution wavelet- transformed image analysis of histological sections of breast carcinomas, CellularOncology, vol. 27, pp. 237-244, 2005.
- [21] Defeng Wang, Lin Shi, Pheng Ann Heng - Automatic detection of breast cancers in mammograms using structured support vector machines, Neuro-computing, Volume 72, issue 13-15, Pages 3296-3302, August 2009.
- [22] Thème de magister en Électroniques biomédicale « *Application à la détection des lésions mammaires* », Hadhidj Ismahen, Tlemcen 2012.
- [23] Mémoire de fin d'études pour l'obtention du diplôme de Master en Informatique « *Segmentation des masses dans les images mammographiques* » Hamza CHEKKAF, Ibrahim TOUIL, Tlemcen 25 septembre 2012.
- [24] Nawel Hamdan, Bellabes zineb, « Analyse de texture des masses mammaires », Laboratoire de Génie Biomédical, Tlemcen, Université de Tlemcen, Algérie, 2014.
- [25] Hanifi Majdoulayne, « *Extraction de caractéristiques de texture pour la classification d'images satellites* », Thèse DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE ,2009.
- [26] User, M., 1984. « Description statistique de la texture, » Thèse de docteur en sciences Ecole Polytechnique Fédérale de lausanne, suisse.
- [27] Maitre H, Le traitement des images, Hermes Science Publications, 364 pages, 2003.
- [28]. S. Bonnevey, Extraction de caractéristiques de texture par codages des extrema de gris et traitement prétopologique des images, PHD, Université Claud Bernard Lyon 1, France, 29 octobre 1997.
- [29] Haralick RM, Statistical and structural approaches to texture, Proceedings of the IEEE, 67(5), pp. 786-804, 1979.
- [30]. F. Eddaoudi, F. Regragui, Characterization of normal mammograms based on the Statistical Features, Proc Second International Symposium on Communications, Control and Signal Processing (ISCCSP), Marrakech, Morocco, Mars , 2006.
- [31] B. Mohammed et M. Benazzouz « *Segmentation d'images microscopiques basée sur les attributs textures* », Laboratoire Génie Biomédical, Tlemcen, 2014.
- [32] Mario Youssef GHARIB , « DETECTION DES CHANGEMENTS POUR LES OBSERVATOIRES PHOTOGRAPHIQUES DU PAYSAGE », UNIVERSITE LIBANAISE ECOLE DOCTORALE DES SCIENCES ET DE TECHNOLOGIE ,2011.
- [33] CCRS, Notions fondamentales de télédétection, Classification et analyse des images [http://www.ccrs.nrcan.gc.ca/resource/tutor/fundam/chapter4/07_f.php]
- [34] A tutorial on clustering algorithm. [http://home.dei.polimi.it/matteucc/Clustering/tutorial_html].
- [35] http://fr.wikipedia.org/wiki/Apprentissage_semi-supervis%C3%A9.

- [36].Kilday, J., Palmieri, F. et Fox, M. (1993). Classifying mammographic lesions using computerized image analysis. *IEEE Transactions on Medical Imaging*, 12(4):664–669.
- [37]. Rangayyan, R., El-Faramawy, N., Desautels, J. et Alim, O. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810.
- [38].Huo, Z., Giger, M. L., Vyborny, C. J., Wolverton, D. E., Schmidt, R. A. et Doi, K. (1998). Automated computerized classification of malignant and benign masses on digitized mammograms. *Academic Radiology*, 5(3):155–168.
- [39].Bruynooghe, M. (2006). Mammographic mass detection using unsupervised clustering in synergy with a parcimonious supervised rule-based classifier. *International Workshop on Digital Mammography (IWDM)*, pages 68–75.
- [40] Viton, J.-L., Rasigni, M., Rasigni, G. et Llebaria, A. (1996). Method for characterizing masses in digital mammograms. *Optical Engineering*, 35(12):3453–3459.
- [41].Kegelmeyer, W., Pruneda, J., Bourland, P. et al. (1994). Computer-aided mammographic screening for spiculated lesions. *Radiology*, 191:331–337.
- [42].Li, L., Mao, F., Qian, W. et Clarke, L. (1997). Wavelet transform for directional feature extraction in medical imaging. *IEEE International Conference on Image Processing*, volume 3, page 500, Los Alamitos, CA, USA. IEEE Computer Society.
- [43]Kim, H. et Kim, W. (2005). Automatic detection of spiculated masses using fractal analysis in digital mammography. *Computer Analysis of Images and Patterns (CAIP)*, pages 256–263.
- [44] Alice POREBSKI, « *Sélection d'attributs de texture couleur pour la classification d'images. Application à l'identification de défauts sur les décors verriers imprimés par sérigraphie.* », THÈSE DOCTEUR DE L'UNIVERSITÉ DE LILLE 1, 2009.
- [45] Imane NEDJAR , « CMBIR (Content Medical Based Image Retrieval) développement d'outil logiciel d'annotation d'images médicales, utilisant les méthodes d'indexation par descripteurs invariants de contenus » , Mémoire de magister En Informatique, Tlemcen, 2012.
- [46]N. Nacereddine*, S. Tabbone*, D. Ziou**, L. Hamami, « un descripteur efficace pour la reconnaissance des symboles graphiques basé sur la transformée de radon » Dpt d'Electronique, Ecole Nationale Polytechnique, Algérie, 2010.
- [47] Reconnaissance de caractères: Descripteurs de formes et classification Philippe Foucher. www2.univ-paris8.fr/ingenierie/masterfoucher/rd3.pdf.
- [48] N.HAMDAN, M.Himmi, I.Hadjidj, A.Chikh, Classification neuronale des masses mammaires, The Biomedical Engineering International Conference BIOMEIC'14, Tlemcen, 2014.
- [49] M. Hasan et F. Boris « SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges », BD Web, ISTY3 Versailles St Quentin, France 2006.

Remerciements :

Avant d'entamer ce projet de fin d'étude, nous remercions d'abord ALLAH le tout puissant de m'avoir donné la volonté pour accomplir ce modeste travail.

J'adresse tout d'abord toute ma gratitude et manifester ma très sincère reconnaissance à Monsieur CHARH MOHAMMED EL AMINE de m'avoir proposé ce sujet de mémoire, et de l'attention qu'il a porté à mon travail. J'ai découvert grâce à lui le monde de la recherche dans les meilleures conditions. J'ai beaucoup apprécié les qualités pédagogiques dont Monsieur CHARH MOHAMMED EL AMINE a fait preuve pendant une année.

Je remercie chaleureusement, Monsieur BECHAR HASSANE professeur à l'Université de Tlemcen, d'avoir accepté de présider ce jury, et qui m'a honoré de sa présence.

Je tiens à remercier Madame BENCHAB WASSIM, doctorante à l'Université de Tlemcen, qui m'a fait l'honneur d'examiner ce travail.

Aussi, Mademoiselle BENOUIS AMEL, HAMDAN NAVAL, et MANE NEDJAR, Monsieur BENOUMAR MOHAMMED EL AMINE et BECHAR AMINE, pour le temps et l'intérêt qu'ils voudront bien consacrer à l'évaluation de ce projet.

Mes sincères salutations s'adressent à toute ceux qui ont su supporter ma joies et ma peines,

et

Zui ont su être la quand on en avait besoin.

Enfin un grand merci à ceux qui de proche ou de loin ont

Participé à l'accomplissement de ce travail.

Dédicace :

A mes glorieux Nation

A mes chers parents pour leur soutien inconditionnel,

A mes frères, mes sœurs,

A mes nièces et neveux,

A mon cousin CHEZLI Benchouhra.

A DJALTI Yousef, KHADRAOUI Mohammed

A ma belle-sœur HANANE BENMAMARE (ربي يرحمها)

A tous mes amis, surtout (Sadia, Khadija, Amina, Saliha, Ikrame, Nouhad), avec lesquels j'ai eu de nombreuses et enrichissantes conversations qui m'ont permis d'élargir l'horizon de cette recherche.

*A tous ceux qui travaillent pour que cette
Nation Soit meilleure.*

Résumé :

Le cancer du sein, forme la première cause de mortalité chez les femmes dans le monde, Aujourd'hui, Il est devenu une urgence pour lutter contre ce type de cancer. Pour cette raison, plusieurs études ont montré la nécessité d'une lecture attentive des clichés mammographiques pour aider les médecins dans le dépistage précoce de ce dernier.

L'objectif est de développer des systèmes puissants pour renforcer la reconnaissance des zones anormales, qui sont connus sous le nom SAD (système d'aide au diagnostic). nous développons des classifieurs pour l'identification des masses mammaires en se basant sur les techniques des RNs (les réseaux de neurones) et le SVM (support à vecteur machine).

Notre but principal est de classer les masses mammaires après la détection et la caractérisation des différents descripteurs contenus à l'intérieur de l'image mammographiques. En s'appuyant sur la matrice de cooccurrence. Nous avons utilisé les informations textuelles contenues dans les tissus mammaires exprimées en niveau de gris, et sur les descripteurs de formes géométriques. Le choix de ces descripteurs constitue un problème majeur dans les systèmes d'analyse d'images, car ces derniers conditionnent fortement le résultat final de la classification. Donc, il s'agit de trouver une caractérisation des images plus riche en information que les images d'origine et ceci dans le but de rendre la prise de décision plus efficace. Nous avons validé nos classifieurs sur un échantillon de 90 images de la base de données DDSM. Les résultats expérimentaux ont montré l'intérêt des descripteurs choisis.

Mots-clefs: Cancer du sein, Diagnostic Assisté par Ordinateur (DAOx), Segmentation, classification, Descripteurs de texture et de forme, SVM, RNs.

Abstract :

Breast cancer, the leading cause of death among women worldwide, today is urgently improve the fight against this cancer. For this reason, several studies proved the necessity of the second reading of mammographical clichés to help doctors in the early detection of the latter.

The objective is to develop powerful systems to reenforce the recognition of abnormal areas, which are known as Computer Assisted diagnosis (CAD). One of these systems will we use in the diagnosis of breast masses is based on the supervised learning technique using SVM and neural networks (RNs). Our main goal is to classify breast masses after the detection and characterization of various descriptors contained within the mammographic image, based on the co-occurrence matrix in order to encode information in textural fabrics Breast expressed in grayscale, and the descriptors of geometric shapes. The choice of these descriptors is a major problem in image analysis systems because these descriptors strongly influence the final result of the research or of the classification. So it's the best characterization of the cabled image which make final decision more effective. We have validated our system on a sample of 90 frames (DDSM) and the experimental results show the effectiveness of these descriptors.

Keys words : Breast cancer, classification, segmentation, SVM, RNs, Feature Extraction..

ملخص:

سرطان الثدي، يمثل السبب الرئيسي للوفاة بين النساء في جميع أنحاء العالم، اليوم وجب على وجه السرعة مكافحة هذا السرطان. لهذا السبب، العديد من الدراسات بينت الحاجة إلى قراءة ثانية شريط التصوير الطبي من أجل مساعدة الأطباء في الكشف المبكر عن هذا الأخير. الهدف هنا هو تطوير أنظمة قوية وفعالة لتحقيق أقصى قدر ممكن في استكشاف المناطق الغير طبيعية، والتي تعرف باسم الانظمة المساعدة (SAD). احد هذه الأنظمة التي سوف نستخدمها في تشخيص أنسجة الثدي يستند بطريقة معينة على استخدام تقنية طريقة التصنيف باستخدام SVM والشبكات العصبونية (RNs). هدفنا الرئيسي هو تصنيف أنسجة الثدي بعد الكشف وتمييز مختلف المحددات الموجودة في صورة الأشعة، بالاعتماد على مصفوفات التساوق بهدف ترميز المعلومات النسيجية المحتواة في أنسجة الثدي الممتلئة بالتدرج الرمادي و واصفات الأشكال الهندسية. اختيار هذه الصفات بشكل مشكلة رئيسية في نظم تحليل الصور لأن هذه الأوصاف تؤثر بشدة على النتيجة النهائية للبحث أو التصنيف. لذلك وجب العثور على ترميز أكثر ثراء بالمعلومات من الصور الأصلية وذلك من أجل جعل عملية اتخاذ القرار أكثر فعالية. لقد جربنا عينة من 90 صورة و النتائج التجريبية تثبت فعالية هذه الوصفات.

الكلمات المفتاحية : سرطان الثدي، الانظمة المساعدة، والشبكات العصبونية، التصنيف، واصفات الأشكال الهندسية، التجزئة

