

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البلدية
Université SAADDAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



MÉMOIRE DE MASTER

Mention Génie Biomédical

Spécialité : Instrumentation Biomédicale

présenté par

AGGOUNI Manelle

&

HADJ AHMED Soumia

le:10Juillet2019

Détection automatique des masses mammaires

Proposé par : BECHAR Mohammed El Amine

Année Universitaire 2018-2019

Remerciements

Avant tout nous remercions "ALLAH" le tout puissant de nous avoir donné le courage, la volonté, la patience et la santé durant toutes ces années et que grâce à lui ce travail a pu être réalisé.

Nos remerciements les plus vifs et chaleureux, vont à notre promoteur de mémoire Mr. BECHAR Mohammed El Amine, pour son aide, son orientation judicieuse et sa disponibilité, aussi pour la confiance, la patience et la compréhension qu'il nous a toujours manifesté.

Nos vifs remerciements vont également aux membres du jury d'avoir accepté de corriger et d'évaluer notre travail.

Nous voulons dire merci à tous les enseignants du département d'électronique de l'université de Blida pour l'aide pendant notre formation d'étude, et à toutes personnes qui ont aidé de proche ou de loin.

Enfin, on ne peut pas oublier de remercier nos familles, en particulier nos parents, pour leur compréhension, leurs sacrifices et leur patience.

Dédicace

A ma très chère mère Quoi que je fasse ou que je dise, je ne serai Paient de remercier comme il se doit ton affection me couvre la bien vaillance me guide et la présence à mes côtés a toujours était ma source de force pour affronter les différents obstacles

A mon très cher père Tu as toujours été à mes côtés pour me soutenir et m'encouragé. Que se travail traduit me gratitude et mon affection. A mes très chères frères Mohamed, Ahmed, Mustapha et mes sœurs Yasmina, Chahrazed, Nedjla Puissent dieu vous donne santé, bonheur, courage, et surtout réussite.

Manelle

Je tien a dédié ce travail qui un rapport de mémoire à ma chère maman pour son amour infini pour son soutien incorporable, pour sa compréhension qui n'a pas d'équivalent avec mes sentiment d'amour et de respect les plus chaleureux à mon chère père a qui je dois dans et tout symbole de courage et de sacrifice sa patience et son aide qui m'ont toujours encouragé et soutenu au cours de la période de mes études je souhaite que ce travail sera un témoignage de ma profonde affection et reconnaissance de sacrifie de mon père .

À mes chères frères Ridha, Ilyes.

À mes chères sœurs Sabiha, Hanane, et surtout Ouarda.

Soumia

Table des matières

Remerciements	
Dédicace.	
Abstract	
Table des matières	
Table des figures	
Liste des tableaux	
Glossaire	
1 Introduction Générale	i
2 Contexte médical	3
2.1 Anatomie du sein	3
2.2 Généralités sur le cancer du sein	4
2.2.1 Types du cancer du sein	4
2.2.2 Facteurs de risque	5
2.2.3 Symptômes du cancer du sein	5
2.3 Anomalies mammaires	6
2.3.1 Macro-calcifications	6
2.3.2 Micro-calcifications	6
2.3.3 Masse	7
2.4 Les techniques d'aides au diagnostic du cancer du sein	8
2.4.1 Mammographie	8
2.4.2 Échographie mammaire	9
2.4.3 IRM mammaire	10
2.5 Conclusion	11
3 Etat de l'art	12
4 Classification super-pixellique	16
4.1 Préparation de l'image Brute	17
4.1.1 Les artéfacts dans une mammographie	17
4.1.2 Muscle pectoral	17
4.2 Amélioration de l'image	18
4.2.1 Modification d'histogramme	18
4.2.2 La morphologie mathématique	22
4.3 SLIC	23
4.3.1 L'algorithme SLIC (Simple Linear Iterative Clustering)	24
4.4 Extraction de variables	25
4.4.1 Les statistiques de premier ordre	25

4.4.2	Matrice cooccurrence	27
4.4.3	Filtre de Gabor	29
4.4.4	Local Binary Pattern	30
4.4.5	Moments de Hu	31
4.5	Classification supervisée	32
4.5.1	Arbre de décision	32
4.5.2	Support Vector Machine (SVM)	34
4.6	Conclusion	34
5	Résultat et discussion	35
5.1	Base de données	36
5.1.1	Description de la base de données	36
5.2	Paramétrages	37
5.3	Résultats et discussions	37
5.3.1	Préparation de l'image	37
5.3.2	Amélioration de l'image	38
5.3.3	SLIC	40
5.3.4	Détécton automatique des masees mammaires	40
5.3.5	Test d'efficacité de la méthode proposée	41
5.3.6	Présentation de l'application "Inter"	42
5.4	Conclusion	43
6	Conclusion	45
	Bibliographie	47

Table des figures

2.1	La glande mammaire [1].	4
2.2	Macro-calsification mammaire	6
2.3	Micro-calsification mammaire	7
2.4	Masse mammaire.	7
2.5	Les différents attributs de la masse.	8
2.6	Mammographie	9
2.7	Image obtenue par la mammographie	9
2.8	Echographie mammaire	10
2.9	Image IRM mammaire	11
4.1	Processus de la classification superpixelique.	16
4.2	Les artéfacts du film mammographique.	18
4.3	Histogramme d'une image en niveau de gris,(a) image en niveau de gris et (b) histogramme de l'image.	19
4.4	Egalisation de l'histogramme, (a) image originale, (b) histogramme de l'image originale, (c) image après égalisation et (d) histogramme de l'image après égalisation.	20
4.5	Image améliorée par égalisation d'histogramme adaptative.	21
4.6	Élément structurant.	22
4.7	Image super-pixillique.	24
4.8	Extraction de caractéristique.	26
4.9	Transformation d'une image en matrice de cooccurrence, (a) image et (b) la matrice de cooccurrence.	27
4.10	Le Code LBP, (1) image en niveau de gris, (2) Multiplication de code binaire et le seuillage, (3) Représentation de LBP.	31
4.11	Arbre de Décision	33
5.1	Schéma résume notre méthode proposée	35
5.2	(a) Image brute et (b) Image Résultante.	37
5.3	Les résultats de chaque technique d'amélioration.	39
5.4	Résultat de l'algorithme SLIC, (a) k=500 super-pixels, (b) k=1500 super-pixels et (c) k=2000 super-pixels.	40
5.5	Résultat comparatif entre SVM et l'arbre de décision	42
5.6	Classification par SVM.	43

Liste des tableaux

5.1	Les classes de la base de données MIAS.	36
5.2	Description de la base de données.	36
5.3	Paramétrage.	37
5.4	Synthese de PSNR.	39
5.5	Accuracy et le F-score de SVM et Arbre de décision.	41

Glossaire

ANN : Artificial Neural Networks
ACR : American College of Radiology
ACC : accuracy
Arb : Arbre de décision
BI-RADS : Breast Imaging Reporting And Data Systems
BTH : Black Top Hat
BHLE : Band Limited Histogram Equalization
CAD : Computer Aide Diagnostic
CCIS : Cancer Canalaire In Situ
CLIS : Cancer Lobulaire In Situ
CCI : Cancer Canalaire Infiltrant
CLI : Cancer Lobulaire Infiltrant
DAO : Diagnostic Assisté par Ordinateur
DDSM : Digital Databas for Screening Mammographic
DRSLE : Distance Regularized Level Set Evolution
ERLM : Ecran Radio Luminescent à Mémoire
EUS-SVM : Ensemble of Under-Sampled SVM
ES : Elément Structurant
EQM : Erreur Quadratique Moyenne
FCM : fuzzy C Means
FN : Faux Négatif
FP : Faux Positif
FPI : faux positif par image
F : Ferméture
fGabor : filtre de Gabor
GLCM : Gray Level Co-occurrence Matrix.
IRM : Imagerie par Résonance Magnétique
LBP : Local Binary Pattern
LPE : ligne de partage des eaux
mini MIAS : mini Mammogram Image Analysis Society
MLP : Multi-Layer perception
m RMR : minimum Redundancy Maximum Relevance
MLO : Médio Latérale Oblique
MC :Matrice de Cooccurrences
O : ouverture
PSNR : Peak Signal to Noise Ratio
R/L : Right/Left
SVM : Support Vector Machine

SVMRFE : SVM Recursive feater Elimination

SRN : SVM RFE with NMIFS filter

SPF : Signed Range Function

SLIC : Simple Linear Iterative Clustering

VN : vrai Négatif

VP : Vrai Positif

WTH : White Top Hat

Chapitre 1

Introduction Générale

Dans la communauté féminine, le cancer du sein est considéré comme le type de cancer le plus fréquent chez la femme âgée entre 40 et 65 ans, et qui représente l'un des principales causes de mortalité. En Algérie, chaque année, environ plus de vingt mille (20.000) nouveaux cas sont décelés avec une moyenne d'âge de 40 ans, dont 95% des cas se présentent à un stade avancé de la pathologie [2]. Ces chiffres indiquent une situation impressionnante de femmes atteintes de cancer du sein décèdent en raison de l'absence de dépistage à temps.

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

Dans ce but, la technique de radiologie la plus efficace est celle qui met en évidence des lésions au niveau du sein, donc c'est bien la technique de la mammographie [2]. Selon les radiologues, un indicateur important du cancer du sein est la présence des calcifications qui apparaissent dans 30% à 50% des cas diagnostiqués par mammographies [3]. La mammographie est une technique à faible dose de rayons X qui ne résulte pas une bonne visualisation de la structure interne du sein. La faible qualité des images mammographies fait l'objet de la difficulté d'analyse des clichés par les médecins radiologistes. Par conséquent, la conception d'un système de détection assistée par ordinateur (CAD) représente un système pour aider les radiologistes dans l'interprétation des mammographies pour le dépistage de masse et de calcification.

L'objectif de notre travail est de développer un système d'aide au diagnostic dédié pour le dépistage du cancer du sein. Le système a pour but de détecter la présence d'une région anormale dans un cliché de mammographie, cette détection est réalisée par des techniques intelligentes.

Notre plan de mémoire se compose de quatre chapitres :

Le premier chapitre est consacré à la définition des notions de base sur l'anatomie du sein ainsi que la description de pathologies mammaires. De plus, un aperçu sur les techniques d'imagerie médicale sera présenté également dans ce chapitre.

Le deuxième chapitre présente un état de l'art sur les méthodes de segmentation et de classification des images mammographies.

Le troisième chapitre est réservé à la présentation de la méthode proposée et la définition de chaque traitement inclus dans la méthode proposée.

Le quatrième chapitre est consacré principalement à l'étude expérimentale. Dans ce chapitre, nous avons exploité la base de données MIAS pour tester la méthode proposée. De plus, une brève présentation de l'interface développée est donnée.

Nous terminons ce mémoire par une conclusion générale et des futures perspectives.

Chapitre 2

Contexte médical

Le cancer du sein est le cancer le plus commun chez la femme, lorsque ce type de cancer est détecté dans un stade très développé, il va se propager vers d'autres parties du corps, c'est pour cette raison qu'il est préférable de le dépister dans un stade précoce.

Dans la première partie de ce chapitre, nous commençons par introduire la notion de l'anatomie du sein, nous abordons le cancer du sein dans la deuxième partie, et nous allons présenter la routine de diagnostic du cancer du sein dans la dernière partie.

2.1 Anatomie du sein

Les seins sont situés au niveau du thorax, en avant du muscle pectoral de chaque côté et s'étendent en hauteur de la clavicule, et en largeur de l'aisselle jusqu'au milieu du sternum. Ils sont symétriques et ne contiennent pas de muscle [4].

Les seins sont constitués d'une glande mammaire, de tissu conjonctif et de tissu graisseux qui sépare les compartiments et qui donne au sein la forme et le volume. Ils arborent en leur centre une région pigmentée appelée aréole mammaire, qui elle-même entoure le mamelon comme il est montré dans la Figure 2.1. La glande mammaire est formée de 15 à 25 compartiments, chacun de ces compartiments est constitué de lobules et de canaux. Ces lobules renferment des unités sécrétoires et des acinis, qui sont responsables de la production de lait. Le lait est acheminé dans les lobules jusqu'au mamelon par des canaux sécréteurs et des canaux galactophores.

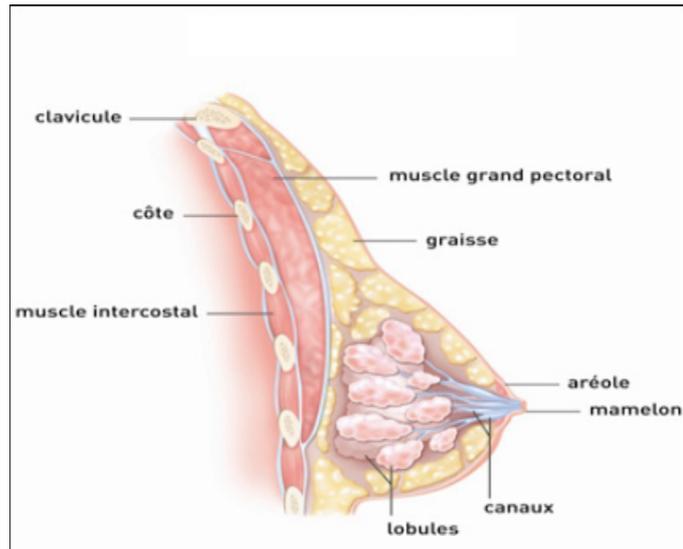


FIGURE 2.1 – La glande mammaire [1].

Le sein est parcouru de vaisseaux sanguins et de vaisseaux lymphatiques [4]. Les ganglions et les vaisseaux lymphatiques composent le système lymphatique qui aide notamment à combattre les infections. Les ganglions lymphatiques du sein sont principalement situés :

- Au niveau de l’aisselle (ganglions axillaires) ;
- Au-dessus de la clavicule (ganglions sus-claviculaires) ; sous la clavicule (ganglions sous-claviculaires ou infra-claviculaires) ;
- A l’intérieur du thorax, autour du sternum (ganglions mammaires internes).

2.2 Généralités sur le cancer du sein

Le cancer du sein est le cancer le plus fréquent chez la femme, il est présenté sous forme d’une tumeur maligne et qui se développe dans le sein.

2.2.1 Types du cancer du sein

Il existe différents types de cancer, les plus fréquents sont les cancers in situ, infiltrant et inflammatoire [4].

Cancer In Situ

Quand les cellules cancéreuses se situent à l’intérieur des canaux ou des lobules, et sans infiltré le tissu voisin, donc on parle du cancer ou carcinome In Situ.

Cancer infiltrant

Quand les cellules cancéreuses ont infiltré le tissu qui entoure les canaux et les lobules, donc on parle de cancer ou carcinome infiltrant, et qui se divise en deux

sous-types :

- **Le cancer canalaire infiltrant** il représente 85% à 90% des cancers du sein infiltrant, et il est défini comme une prolifération de cellules malignes dans le réseau galactophorique.
- **Le cancer lobulaire infiltrant** est le cancer le plus rare, il représente 10 à 15% des cancers du sein lobulaire, et il est considéré comme un facteur de risque qui peut développer un cancer du sein et non comme un précurseur direct de cancer.

Cancer inflammatoire

Les cancers inflammatoires du sein sont des cancers à progression rapide et agressifs. Ils se présentent comme une infection. La rougeur, le gonflement du sein et parfois la peau d'orange s'étendent sur au moins le tiers de la surface du sein et ces signes ont tendance à couvrir rapidement le reste du sein.

2.2.2 Facteurs de risque

Comme le cancer du sein est une maladie multifactorielle, donc il existe plusieurs facteurs de risques [4], qui sont :

L'âge

Le risque d'avoir un cancer du sein augmente avec l'âge même s'il peut atteindre des femmes à des âges très différents [4]. Les femmes de 30 ans ont le risque d'avoir un cancer du sein environ 10%, pour les femmes de 50 à 69 ans environ 50%, et pour les femmes de plus de 70 ans environ 28%.

L'antécédent familial

Près de 20 à 30% des cancers du sein se manifestent chez des femmes ayant des antécédents familiaux. D'après des études scientifiques [4], ils ont constaté que :

- Si la parente au premier degré (la mère, la sœur ou la fille) qui a déjà eu un cancer du sein, le risque d'être atteint de cancer est environ deux fois plus élevé.
- Si la parente au second degré (comme une grand-mère, une tante ou une nièce, d'un côté ou l'autre de la famille) qui ont déjà été atteintes du cancer du sein, ce risque augmente légèrement.

2.2.3 Symptômes du cancer du sein

Il existe plusieurs symptômes du cancer du sein [4], parmi ces symptômes nous pouvons citer :

- L'apparition d'une grosseur (boule) au niveau du sein ou de l'aisselle ;
- Une déformation du sein ;
- Une modification de la texture ou de pigmentation de la peau du sein ou de l'aréole ;
- L'apparition d'écoulements anormaux au niveau du mamelon.

2.3 Anomalies mammaires

La généralisation du dépistage du cancer du sein, qu'il soit du dépistage individuel ou du dépistage de masse organisé, amène à découvrir beaucoup d'anomalies purement radiologiques [4]. Parmi ces anomalies, on trouve :

- Macro-calcification.
- Micro-calcification.
- Masse.

2.3.1 Macro-calcifications

Ce sont des gros dépôts calcaires dans le sein qui sont associés à des affections bénignes comme par exemple une ancienne plaie, infection ou un kyste calcifié. Figure 2.2 montre les Macro-calcification dans une image mammaire de la base de données MIAS de référence mdb186 .

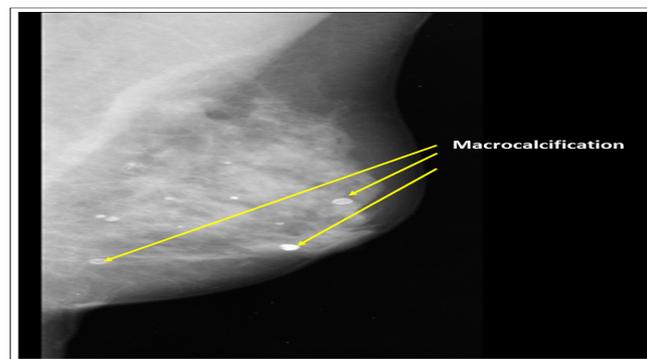


FIGURE 2.2 – Macro-calsification mammaire

2.3.2 Micro-calcifications

Ce sont des dépôts calcaires minuscules ($< 0,5$ mm) qui ressemblent à des grains de sable dans le sein, comme il est illustré dans la Figure 2.3 correspond a l'image mdb148 de la base de données MIAS. Elles sont facilement vues en mammographie du fait de leur forte densité (elles apparaissent comme des ponctuations blanches sur le gris des tissus). Elles représentent la majorité des problèmes rencontrés sur les mammographies de dépistage. Quand elles sont nombreuses, de forme irrégulière et groupées, cela peut être un signe de cancer du sein. Dans ce cas une biopsie est indiquée.

Le seul signe précoce de cancer que l'on peut observer sur une mammographie est la présence de micro-calcifications. Dans 40 à 70% des cas elles sont associées à un processus malin.

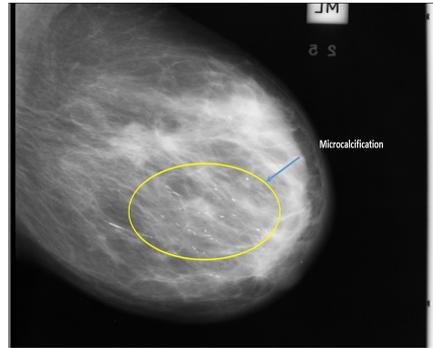


FIGURE 2.3 – Micro-calsification mammaire

2.3.3 Masse

Une masse est une opacité importante occupant un espace dans le sein et vue comme une tâche blanche sur l'image mammographique (voir la Figure 2.4). Il peut s'agir d'un kyste (collection liquidienne non cancéreuse), d'une lésion solide ou un cancer du sein. La masse peut être caractérisée par des attributs, qui permettent aux médecins de la décrire en vue de déterminer sa nature, sa forme, son contour et sa densité (Voir la Figure 2.5 correspond à l' image mdb025 de base de données MIAS).



FIGURE 2.4 – Masse mammaire.

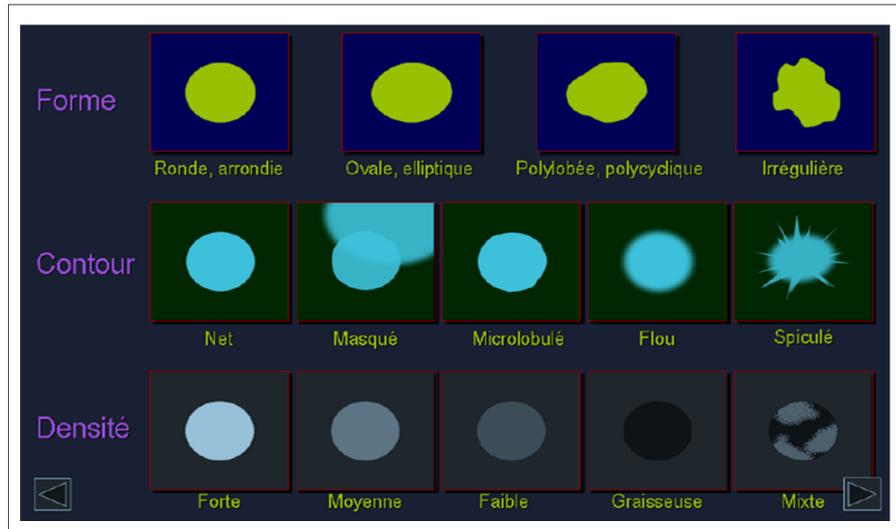


FIGURE 2.5 – Les différents attributs de la masse.

2.4 Les techniques daides au diagnostic du cancer du sein

2.4.1 Mammographie

La Mammographie [5] est un examen radiologique consacré à la détection des pathologies du sein (Voir la Figure 2.6). C'est la plus performante en matière de surveillance et de dépistage précoce du cancer du sein. La qualité de l'image mammographie est caractérisée par trois paramètres physique : **la résolution, le contraste et le bruit** .

La mammographie est caractérisée par deux types ; une mammographie **conventionnelle** (analogique, ou « classique ») qui produit une image imprimée sur un film argentique (un cliché), où une mammographie **numérique** qui construit l'image numérique à partir de la projection de rayons X.



FIGURE 2.6 – Mammographie

Durant l'examen, la patiente se tient debout devant un appareil de radiologie, torse nu. Chacun de ses seins est comprimé entre deux plaques pour prendre deux clichés, de face et de profil. L'examen dure une dizaine de minutes au total. Les images obtenues sont analysées par un radiologue qui y recherchera la présence d'anomalies. La figure (2.7) illustre une image acquise par un mammographe.



FIGURE 2.7 – Image obtenue par la mammographie

2.4.2 Échographie mammaire

Une échographie [5] est très utile pour identifier des liquides et analyser la structure des nodules palpés ou découverts sur une mammographie (Voir la Figure 2.8). Une échographie complète donc certaines mammographies difficiles à interpréter, mais elle ne peut jamais remplacer une mammographie car une échographie n'analyse pas certaines structures, sauf dans des cas particuliers : femmes enceintes ou adolescentes. Le médecin examine chacun des seins à l'aide d'une sonde après avoir appliqué un gel sur la peau. L'examen est totalement indolore.



FIGURE 2.8 – Echographie mammaire

2.4.3 IRM mammaire

L'IRM [5] du sein ne remplace pas la mammographie ni l'échographie mammaire, la Figure 2.9 représente une image IRM. Elle ne constitue pas un examen systématique pour le diagnostic du cancer du sein, c'est une technique non traumatisante et non irradiante. Il s'agit d'un outil supplémentaire qu'on emploie parfois :

- Pour faire la différence entre une anomalie bénigne et une anomalie cancéreuse dans le cadre du dépistage chez les femmes qui risquent de contracter un cancer du sein ;
- Lorsque l'imagerie standard (mammographie ou échographie) ne permet pas de conclure avec certitude à l'absence de malignité ;
- Pour guider un prélèvement par biopsie ; dans le cadre du bilan d'extension ;
- Pour évaluer la réponse aux thérapeutiques néo adjuvantes (chimiothérapie réalisée avant l'opération) ;
- A la recherche de récurrence locale après la chirurgie conservatrice ;
- Pour vérifier l'état d'implants mammaires, en particulier si on croit qu'il y a eu rupture.

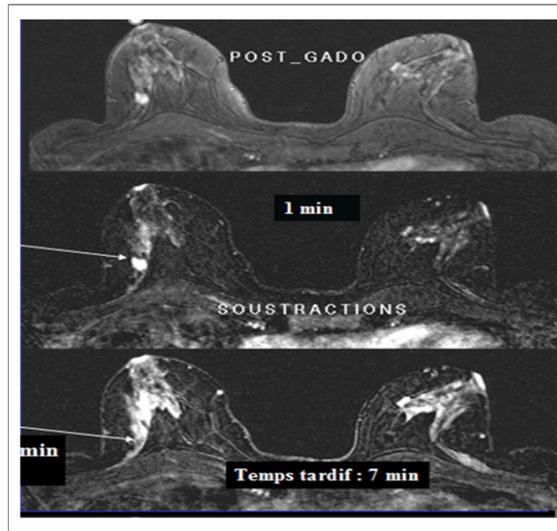


FIGURE 2.9 – Image IRM mammaire

2.5 Conclusion

Dans ce chapitre, la notion du cancer du sein ainsi que la mammographie ont été présentés. Une grande attention a été consacrée à l'étude des spécifications des pathologies mammaires à savoir les masses mammaires. Le chapitre suivant sera consacré à étudier les travaux antérieurs permettant la conception des systèmes d'aide au diagnostic pour le dépistage du cancer du sein.

Chapitre 3

Etat de l'art

Les images mammographiques occupent une place primordiale pour la détection des tumeurs mammaires, elles fournissent des informations radiologiques tels que la nature, le type et l'état de la tumeur, et si on arrive à détecter toutes ces informations nous améliorons le pronostic vital des patientes. A ce titre, plusieurs travaux ont été effectués afin de développer des outils d'aide au diagnostic de cette maladie cancéreuse. La mammographie est une technique d'imagerie médicale qui permet d'obtenir une image représentant la structure du sein. Il existe plusieurs travaux sur la classification des images mammographiques.

- El Atlas et al [6] ont proposé une nouvelle technique de caractérisation basée sur la fusion d'information (fusion des caractéristiques). Les caractéristiques sont extraites à partir de la morphologie de la masse mammaire pour être utilisées dans le processus de la fusion. Ce processus comprend quatre étapes : segmentation, calcul et fusion des caractéristiques morphologiques et classification en utilisant le classifieur Support Vectors Machines (SVM) et Artificial Neural Networks (ANN). Ils ont utilisé INBREAST comme une base de données, les résultats obtenus favorisent la méthode d'ANN pour identifier les masses malignes.
- A.Ibrahim et al [7] ont présenté un algorithme qui aide les radiologues à identifier les tumeurs mammaires à leurs débuts. Cet algorithme élimine automatiquement les artéfacts et le bruit présent dans l'image en utilisant des opérations morphologiques, puis le contraste de l'image sera amélioré en utilisant la méthode BLHE (Band Limited Histogram Equalization). La détection de la région d'intérêt (tumeur) est réalisée en utilisant la méthode de seuillage d'Otsu.
- Liu et Tang [8] ont étudié la classification des masses mammaires. Dans une première étape, ils ont intégré la méthode FCM (Fuzzy C Means) pour la segmentation. Dans la seconde, chaque masse était représentée en caractéristique de géométrie et de texture. La tâche de la classification a été effectuée par LDA+KNN (LDA : Linear Discriminant Analysis et KNN : K-Nearest Neighbors) et le SVM (Support Vectors Machines) avec le noyau gaussien. Le score F, le relief, les SVMRFE (SVM Recursive feature Elimination), mRMR (minimum Redundancy Maximum Relevance) et le SRN

(SVM RFE with au NMIFS filter) sont les méthodes de sélection des caractéristiques utilisées. La performance obtenue par la méthode proposée donne des meilleurs résultats avec une précision 0.9615.

- Hang et Sohn [9] ont proposé une nouvelle méthode pour la segmentation des régions d'intérêt dans les images de mammographie, particulièrement les masses mammaires, muscle pectorale et les régions denses. Ils ont proposé une représentation topographique appelée carte isocontour (une région saillante forme un motif de contour dense quasi concentrique). L'arbre d'inclusion sert à analyser la structure topologique et géométrique de l'image. La saillante d'une région est mesurée topologiquement comme la profondeur minimale de nidification. Les caractéristiques à différentes échelles sont analysées dans des cartes isocontours multi échelle. Cette méthode peut servir comme un système de détection de mammographie.
- Hu et al [10] ont proposé un algorithme de détection qui utilise la combinaison d'une segmentation adaptative globale et locale par seuillage sur une représentation multirésolution de la mammographie. Ils ont utilisé la mini MIAS (mini Mammogram Image Analysis Society) comme base de données et ils ont calculé 2 transformations en ondelettes pour détecter des lésions suspectes. Les étapes proposées sont comme suit : un prétraitement de l'image originale, qui consiste à effectuer une amélioration par un filtre morphologique, une segmentation par un seuillage global, une segmentation locale adaptative basée sur des fenêtres petites et grandes, et enfin un post-traitement à la fin de la segmentation.
- Ferkous et Marouani [11] ont proposé un système de diagnostic assisté par ordinateur (DAO). Ils ont utilisé Multi-Layer perception (MLP) pour classer la mammographie en quatre classes. Les auteurs ont suivi la classification BI-RADS (Breast Imaging Reporting and Data Systems) dans leurs expérimentations, les classes sont : Sein grasseux homogène (Bi-RADS 0), Sein grasseux hétérogène (Bi-RADS 3), Sein dense hétérogène (Bi-RADS 4) et Sein dense homogène (Bi-RADS 5). Par la suite, ils ont classé les images mammographies en malignes ou bénignes.
- Shafiullah et Kwang Nam [12] ont proposé une nouvelle méthode de segmentation (contour active) pour sélectionner les régions d'intérêts dans une image de mammographie en utilisant la base de données mini MIAS (mini Mammogram Image Analysis Society). Ce processus est basé sur une formulation de niveau bimodale, qui détecte les régions de haute intensité dans la mammographie en se basant sur la fonction SPF (Signed Range Function) (La fonction a la valeur dans la plage de $[-1,1]$ à l'intérieur et à l'extérieur du contour).
- Shi et al [13] ont proposé un pipeline automatique pour les images mammographies basé sur la segmentation et la détection mammaire. Le processus peut estimer la limite peau-air à l'aide d'une carte de gradient, détecter la

limite pectorale-poitrine par un étiquetage pixel par pixel, et détecter les calcifications à l'intérieur du sein par un nouveau filtre de texture.

- Martins et al [14] ont proposé un algorithme de détection des masses mammographies pour le dépistage du cancer du sein. Les auteurs ont utilisé les caractéristiques de la matrice de cooccurrence et l'algorithme de K-means pour effectuer un regroupement des structures du sein dans la mammographie. Puis une classification des structures segmentées en masse ou non-masse est réalisée par les machines à vecteurs de support (SVM). Les résultats de cet algorithme montrent une précision de 85%.
- Ren et Jinchang [15] ont réalisé une étude sur le SVM (support vector machine) et l'ANN (artificial neural network) pour analyser le meilleur classifieur des masses mammaires. Ils ont utilisé la base de données DDSM (Digital Database for Screening Mammography) des leurs expérimentations. Ils ont constaté que le SVM utilise un mappage non linéaire pour séparer les données en classes, par contre l'ANN utilise une connexion multicouche et diverses fonctions d'activation pour traiter les problèmes de classification.
- Rahli et benamrane [16] ont présenté une nouvelle approche hybride pour la détection des lésions mammaires basée sur la méthode de K-mean et la méthode de search harmony. L'approche utilise l'algorithme support vecteur machine (SVM) pour la détection des masses mammographies qui donne une efficacité dans le dépistage du cancer du sein.
- Djebar et al. [17] ont proposé une nouvelle méthode de classification basée sur le classifieur bayésien Naïf et les arbres de décision. La construction d'un arbre de décision se base sur la partition récursive de l'ensemble d'apprentissage en sous-ensembles plus homogènes que l'ensemble de la base. Les deux classifieurs aident à la classification des tumeurs du sein en utilisant des symptômes cliniques, et des caractéristiques extraites à partir de l'image de mammographie. MIAS est la base de données utilisée dans leurs expérimentations et les résultats obtenus sont satisfaisants.
- Chu et al [18] ont proposé une méthode de détection de masse mammaire pour un système DAO (Diagnostic Assisté par Ordinateur). La méthode est basée sur une opération morphologique pour l'élimination des objets plus gros que la masse, et l'élimination des objets plus petits que la masse. La méthode de SLIC (Simple Linear Iterative Clustering) est utilisée pour diviser l'image en sous-région (super-pixel), qui utilise les super-pixels. Puis, la présélection des super-pixels est réalisée par un classifieur de type EUS-SVM (Ensemble of Under-Sampled SVMs). Un raffinement des contours est réalisé par la méthode DRSLE (Distance Regularized Level Set Evolution) pour réduire le taux de FP (Faux Positif). Ils ont utilisé DDSM (Digital Database for Screening Mammography) comme une base de données. Les résultats obtenus ont préservé plus le détail des contours de masses.

- Guarroudji [19] est intéressée à l'étude de la segmentation des clichés mammographies en vue de la détection de foyer de micro-calcifications mammaires. Pour cela, il a commencé par améliorer les images en utilisant un filtre gaussien puis prétraiter les images en utilisant la morphologie mathématique (chapeau haut de forme blanc et chapeau haut de forme noir). Ensuite, il a segmenté les images par la méthode d'Otsu et la méthode de contour actif.
- Chabane et Haddache [20] sont intéressés à la segmentation des images de mammographie. Pour cela, ils sont passés par toute la chaîne de traitement d'image, le prétraitement en utilisant un filtre médian, puis la segmentation par deux algorithmes LPE (ligne de partage des eaux) et le contour actif. Ils sont passés par la suite à la classification et la reconnaissance de forme en utilisant les critères de circularité pour automatiser la classification des masses (maligne/bégnine).
- Hongyu et al [21] ont proposé un nouveau cadre pour détecter les masses mammaires dans les mammographies numérisées. Leur algorithme peut être divisé en trois modules : l'intégration de sensation qui est le sens visuel du radiologue qui joue un rôle important dans la détection de masse, consistant à sélectionner et à obtenir des informations visuelles à partir de pixels de l'image, mais il reste encore beaucoup de patches faux positifs. Donc ils ont introduit l'intégration sémantique afin de mieux distinguer les patches de masses du tissu normal en fonction des caractéristiques morphologiques du tissu mammaire, pour aider à séparer la masse du tissu normal. Pour la dernière étape, la vérification qu'est la réduction des faux positifs basée sur la classification des caractéristiques de texture. Ce cadre peut être considéré comme une méthode d'échange réciproque, il s'agit d'une nouvelle méthode d'exploration pour la détection automatique des lésions. Deux bases de données ont été utilisées (MIAS) et (DDSM).

Dans ce chapitre, nous avons évoqué un ensemble des recherches sur notre sujet, qui peuvent orienter notre choix pour proposer une approche qui permet d'aider le médecin dans le diagnostic.

Chapitre 4

Classification super-pixellique

Le traitement d'image est un processus qui commence par le prétraitement et qui se termine par la classification. Dans ce travail, nous nous intéressons à la classification par super-pixellique. C'est une technique de segmentation qui se base sur la classification des super-pixels. Ce dernier est défini comme un groupe de pixels voisins et homogènes.

Comme il a été décrit dans la littérature, la classification super-pixellique [22], c'est celle qui se base sur la segmentation par SLIC (Simple Linear Iterative Clustering). C'est tout un processus enchainé qui commence par la préparation et l'amélioration de l'image (prétraitement), génération de super-pixel, extraction de variables, et finalement une classification pour la segmentation et détection automatique de la masse, la Figure 4.1 illustre le processus de la classification super-pixellique.

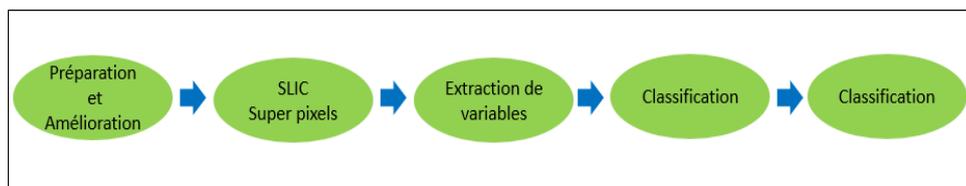


FIGURE 4.1 – Processus de la classification superpixellique.

Le processus est présenté comme suit :

- **Préparation de l'image** : C'est la première étape de ce processus, faite pour éliminer le muscle pectorale et l'étiquette de l'image par le seuillage, et aussi l'élimination des rai numériques par un filtre surfacique.
- **Amélioration de l'image** : La 2eme étape de ce processus, est pour l'élimination de bruit par l'égalisation adaptative de l'histogramme et les opérations de la morphologie mathématique.
- **SLIC** : Segmentation de l'image en régions par superpixels.

- **Extraction de variables** : La caractérisation de chaque région segmentée.
- **Classification** : Classification de la masse mammaire cancéreuse/saine.
- **Segmentation** : la localisation de la masse mammaire dans l'image .

4.1 Préparation de l'image Brute

La préparation de l'image est la mise en œuvre des processus physiques de formation des images, suivis d'une mise en forme pour que ces images puissent être traitées par des systèmes informatiques. C'est un ensemble d'opérations relatives à la collecte, à l'enregistrement, à l'élaboration, à la modification, à l'édition, ... de données. Le principe général de la préparation de l'image est donc à quelques détails près, un système qui reçoit des images, y applique un traitement et produit une information de nature liée à l'application visée.

Le but de cette étape est d'extraire la région d'intérêt « le sein », en supprimant tous types d'artéfacts. Ali Ibrahim et al [7] ont présenté un algorithme qui élimine automatiquement les artéfacts, et qui donne une étiquette pour chaque objet dans l'image binaire. Notre algorithme se fonde sur un seuil bien choisi pour faire la distinction entre la région d'intérêt, les artéfacts et le muscle pectoral. La région d'intérêt sera récupérée après un seuillage suivi d'une opération du filtrage surfacique, ce dernier a pour but de supprimer toutes informations inutiles.

4.1.1 Les artéfacts dans une mammographie

Une mammographie numérisée renferme généralement deux régions distinctes : la région **exposée du sein** et la région **non exposée** constituant le fond, ce fond peut contenir des artéfacts que le système visuel humain peut facilement ignorer pendant l'interprétation. Par contre un système automatisé doit au préalable identifier et classifier ces artéfacts qui engendrent des erreurs d'interprétation lors de l'analyse de l'images.

Les artéfacts dans la mammographie sont généralement marqués par des étiquettes, ils représentent des indicateurs radiopaques manifestant la latéralité de la mammographie (R/L ou Right/Left) ainsi que des indicateurs de projection MLO/CC (La Médiolatérale Oblique/Craniocaudale). Il existe deux types d'artéfacts radiopaques : les bandes de haute intensité, et les marqueurs opaques. Et également les raies de numérisation qui représentent un extremum local de haute intensité (claire ou sombre). La Figure 4.2 illustre les différents artéfacts qui existe dans la mammographie.

4.1.2 Muscle pectoral

Les images mammaires sont difficiles à interpréter, à cause de la non uniformité de leur fond et la présence d'artéfacts qui peut engendrer un échec lors de la

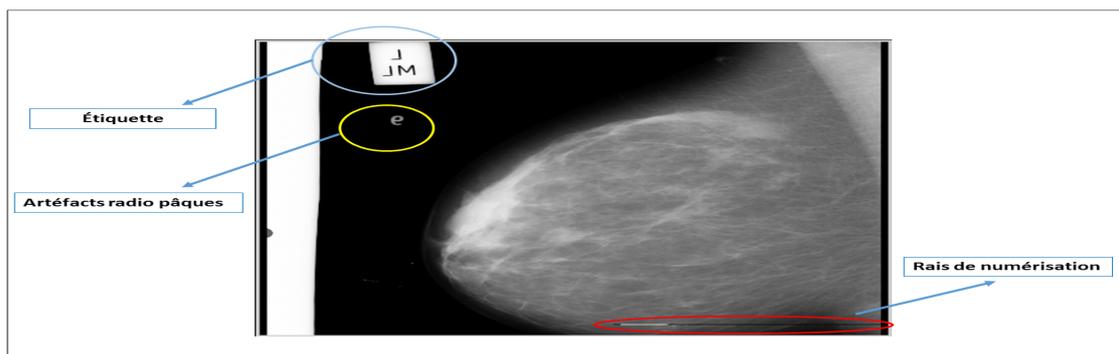


FIGURE 4.2 – Les artéfacts du film mammographique.

segmentation. Pour dégrader cette complexité, une étape nécessaire dans la préparation de l'image qui a pour but d'éliminer le muscle pectoral, et extraire la région du sein.

Avant d'aborder la phase de détection des lésions mammaires sur une mammographie, nous devons passer par deux étapes essentielles, à savoir la détection précise de la région du sein et l'élimination du muscle pectoral. Donc nous proposons de réaliser un seuillage pour détecter respectivement le muscle pectoral et la région du sein. L'extraction de la région du sein permet d'isoler la glande mammaire. Cette étape vise à éliminer le muscle pectoral qui provoque un nombre élevé de faux positifs lors de la détection des masses mammaires.

4.2 Amélioration de l'image

L'amélioration de l'image est une étape indispensable dans la chaîne de traitement d'image, son but est d'améliorer la qualité de l'image lorsqu'elle possède du bruit ou des défauts de contraste. En effet, les structures que l'on recherche n'étant pas toujours facilement discernables, une étape d'amélioration destinée à les mettre en évidence pour faciliter leur détection. Parmi les techniques d'amélioration les plus courantes, nous citons : La modification d'histogramme et la morphologie mathématique. Ali Ibrahim et al [7] ont utilisé un algorithme pour éliminer tous types de bruit sur une image par la morphologie mathématique et l'égalisation d'histogramme.

4.2.1 Modification d'histogramme

Un histogramme [19] est une fonction qui donne la fréquence d'apparition de chaque niveau de gris dans l'image. Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme [7]. Celui-ci permet de donner un grand nombre d'information sur la distribution des niveaux de gris, et de voir entre quelles bornes sont réparties la majorité des niveaux de gris dans les cas d'une image trop claire ou d'une image

trop sombre.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications pour pouvoir extraire les informations utiles de celle-ci. La Figure 4.3 montre un exemple d'un histogramme d'une image en niveau de gris.

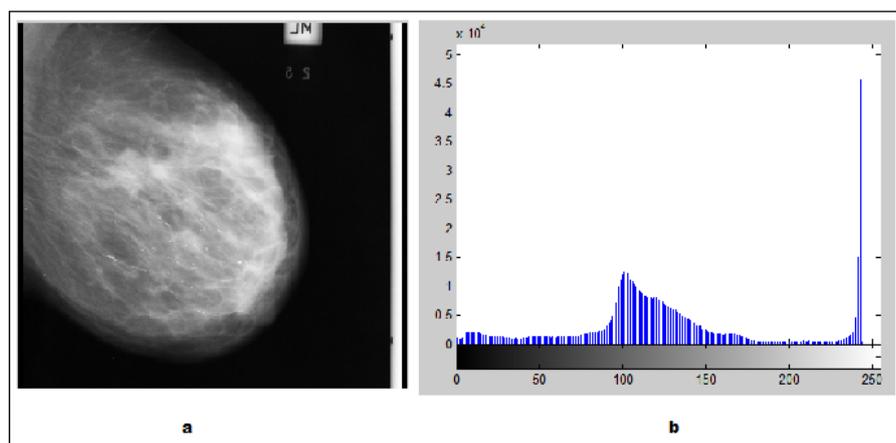


FIGURE 4.3 – Histogramme d'une image en niveau de gris,(a) image en niveau de gris et (b) histogramme de l'image.

Une approche consiste à appliquer une fonction qui associe à chaque valeur d'intensité dans l'image une nouvelle valeur pour modifier les caractéristiques de l'image (accentuer les contrastes en général).

Egalisation de l'histogramme

C'est un outil qui se voit souvent utile pour améliorer certaines images de mauvaise qualité (mauvais contraste, image trop sombre ou trop claire, mauvaise répartition des niveaux d'intensité, etc.). Cette transformation consiste à rendre l'histogramme le plus plat possible. On désire que chaque niveau de gris soit équitablement représenté comme illustré dans la Figure 4.4.

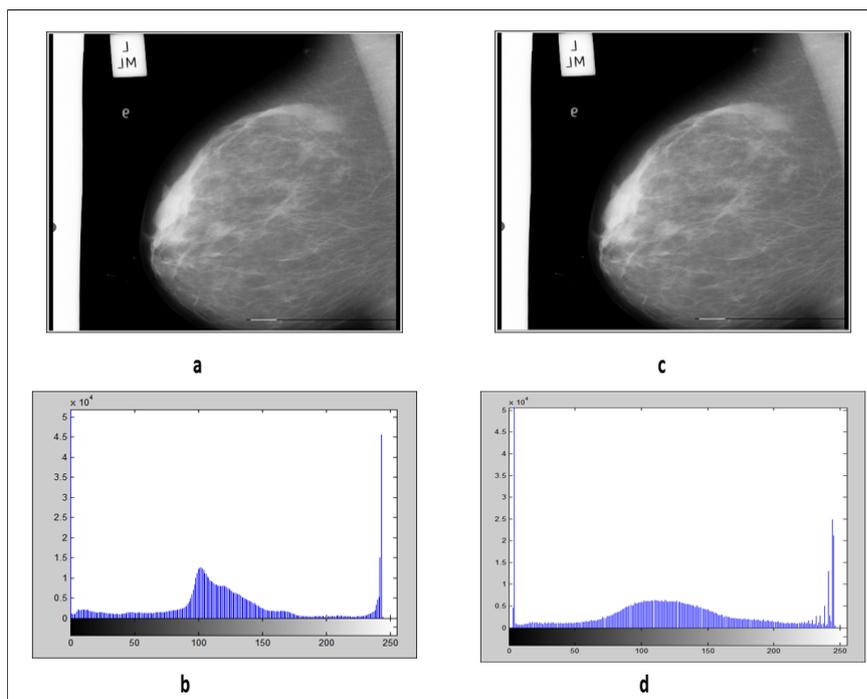


FIGURE 4.4 – Egalisation de l’histogramme, (a) image originale, (b) histogramme de l’image originale, (c) image après égalisation et (d) histogramme de l’image après égalisation.

Egalisation adaptative d’histogramme

C’est une technique de rehaussement d’image. Son principe consiste à appliquer sur chaque pixel ainsi que sa région contextuelle une égalisation d’histogramme (Algorithme 1). Cette région représente en effet les pixels voisins entourant le pixel traité. Le but de cette transformation globale est de distribuer uniformément les tons de gris sur la plage dynamique de l’image, afin d’accroître son contraste. La Figure 4.5 représente une égalisation adaptative d’histogramme.

Algorithm 1 Egalisation adaptative d'histogramme

```

1: Entrée
   moyenne = nombre de pixels / N
   bande = Nr/N
   centre = bande/2
   idebut = 0
   ifin = 0
2: Tant que ifin < Nr faire :
   cumul = 0
3: Tant Que (cumul non proche de moyenne) et (ifin < Nr) faire :
   cumul = cumul + histogramme(ifin)
4: incrémenter ifin
5: Fin Tant Que
6: Pour i = idebut à ifin-1 faire :
   Transfo(i) = centre
7: fin Pour
8: Entrée
   idbut = ifin
   centre = centre + bande
9: fin Tant Que
10: Pour tous les points faire :
     $I(i, j) = \text{Transfo}(I(i, j))$ 
11: fin Pour

```

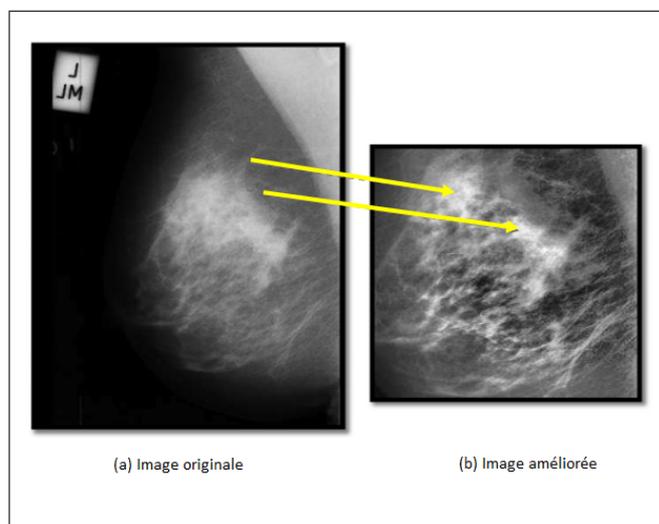


FIGURE 4.5 – Image améliorée par égalisation d'histogramme adaptative.

Avec : Nr = Nombre de pixels de l'image et N : nombre de niveau de gris.

4.2.2 La morphologie mathématique

La morphologie mathématique [19] est une science de la forme et de la structure. Le principe de base de la morphologie mathématique est de comparer les objets que l'on veut analyser à un autre objet de référence, de taille et de forme connues, appelé élément structurant (ES). Ce dernier est un masque de forme quelconque dont les éléments forment un motif, comme dans la Figure 4.6.

En quelque sorte, chaque élément structurant fait apparaître l'objet sous une nouvelle forme. La morphologie consiste à parcourir l'image pour :

- En chaque pixel de départ ($u(i,j)$), on place l'élément structurant centré sur ce le pixel.
- Un test est réalisé pour déterminer la couleur du pixel d'arrivée $u'(i,j)$.

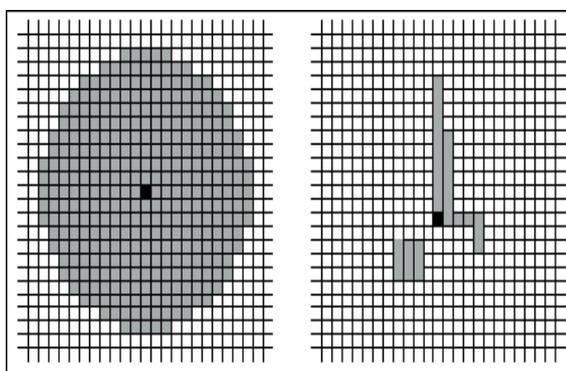


FIGURE 4.6 – Élément structurant.

Cette théorie est fondée sur les opérations d'érosion et de dilatation. Ces opérations de bases, permettent l'ouverture et la fermeture.

Dilatation

C'est la relation imposée entre l'inclusion de la forme étudiée et de l'élément structurant, elle est représentée par la relation mathématique suivante :

$$I \oplus SE = \{z | (SE)_z \cup I \neq \emptyset\}. \quad (4.1)$$

La dilatation va parcourir les pixels de l'image $I(i,j)$ par les étapes suivantes :

- Center l'élément structurant sur ce pixel ;
- Considérer les voisins du pixel dans l'élément structurant (gris+noir) .
- Si l'un de ces pixels est en blanc, mettre $I(i,j)$ en blanc.

Erosion

C'est la relation imposée entre l'intersection de la forme étudiée et l'élément structurant. Sa relation algébrique est :

$$I \ominus SE = \{z | (SE)_z \cap I \neq \emptyset\}. \quad (4.2)$$

L'érosion va parcourir les pixels de l'image $I(i,j)$ par les étapes suivantes :

- Center l'élément structurant sur ce pixel ;
- Considérez les voisins du pixel dans l'élément structurant (gris+noir de l'élément structurant) ;
- Si l'un de ces pixels est en noir, mettre $I'(i,j)$ en noir.

Ouverture

C'est une composition d'une érosion suivie d'une dilatation avec le même élément structurant. Sa relation mathématique est la suivante :

$$I \circ SE = \langle I \ominus SE \rangle \oplus SE = \cup \{(SE)_z | (SE)_z \subseteq I\}. \quad (4.3)$$

Fermeture

C'est une composition d'une dilatation suivie d'une érosion, avec le même élément structurant. Sa relation mathématique est exprimé par :

$$I \cdot SE = \langle I \oplus SE \rangle \ominus SE = \{z | (SE)_z \cap I \neq \emptyset\}. \quad (4.4)$$

Transformation chapeau haut de forme

C'est une transformation qui permet de retrouver les informations éliminées par l'ouverture morphologique ou la fermeture morphologique.

On définit deux types de transformation, la transformation du chapeau haut de forme blanc (WTH) (white top hat, en anglais) qui est définie comme la différence algébrique entre l'identité de l'image (f) et son ouverture $O_B(f)$ tel que :

$$WTH_B(f) = f - O_B(f). \quad (4.5)$$

Par symétrie, pour extraire les Vallès ou ressortir les structures sombres de l'image, on définit le chapeau haute forme noir, (black top hat, en anglais), sa relation algébrique entre la fermeture $F_B(f)$ et l'identité de l'image (f) tel que :

$$BTH_B(f) = F_B(f) - f. \quad (4.6)$$

4.3 SLIC

L'objectif principal du processus de segmentation par SLIC est le partitionnement d'une image en régions d'intérêt du point de vue d'un certain nombre de caractéristiques ou critères [19]. Le SLIC (simple linear itérative clustering) peut segmenter l'image en K régions super-pixels en fonction du contenu de

l'image [23], la segmentation en super-pixel est très utile pour faire de la détection d'objets, pour découper des images et pour alléger la quantité de données à traiter. Segmenter en superpixels permet d'avoir un ensemble de régions d'intérêts à traiter, sans pour autant diminuer la quantité d'information brute dans l'image. La Figure 4.7 représente une image super-pixellique.

Le SLIC est caractérisé par deux paramètres k et m . Le paramètre k est la quantité de super-pixels à générer, l'augmentation de k revient à diminuer la taille des super-pixels dans l'image. Le paramètre m correspond au poids attribués à la distance spatiale par rapport à la distance en couleur Lab . Plus m est fort, moins les couleurs auront un impact lors de la segmentation.

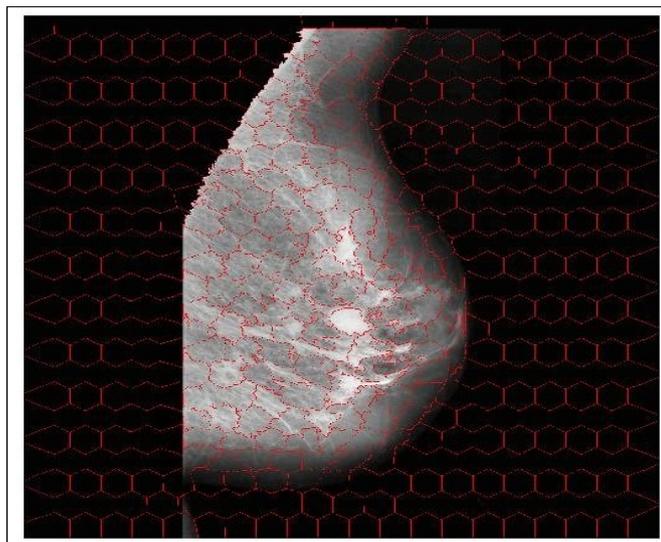


FIGURE 4.7 – Image super-pixillique.

4.3.1 L'algorithme SLIC (Simple Linear Iterative Clustering)

L'algorithme SLIC [23] génère des super-pixels en regroupant les pixels en fonction de leur similarité de couleur et de leur proximité dans le plan de l'image. Ceci est fait dans le plan $[labxy]$ à cinq dimensions, où $[lab]^T$ est le vecteur de couleur de pixel dans l'espace de couleur CIELAB, qui est largement considéré comme perceptuellement uniforme pour les petites distances de couleur, et $[xy]^T$ représente la position du pixel dans le plan de l'image. Alors que la distance maximale possible entre deux couleurs dans l'espace CIELAB (en supposant que les images d'entrée RGB) sont limitées, la distance spatiale dans le plan dépend de la taille de l'image.

L'algorithme SLIC peut regrouper les étapes suivantes :

- Initialisation des centres de clusters. Les centres sont équidistants (distance $S = S = \sqrt{\frac{K}{N}}$), avec N le nombre de pixels dans l'image et K le nombre souhaité de superpixels détaillés.

- Pour chaque pixel i , On associe au centre du cluster le pixel p le plus proche dans la région de recherche, ils se trouvent dans une région de taille $2S \times 2S$ autour du super-pixel.
- Les positions des centres (5 dimensions $[x, y, L, a, b]$ pour leurs coordonnées spatiales et la couleur du pixel dans l'espace LAB), des super-pixels sont mises à jour selon la valeur moyenne de tous les pixels du super-pixel.

$$D_s(C_k, p) = d_{Lab}(C_k, p) + \frac{m}{s} * d_{xy}(C_k, p). \quad (4.7)$$

Où C_k le centre et p le pixel courant.

d_{Lab} est la distance colorimétrique et d_{xy} est la différence entre les positions dans l'image courante, telles que :

$$d_{Lab}(C_k, p) = \sqrt{(L_k - L_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2}. \quad (4.8)$$

$$d_{xy}(C_k, p) = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}. \quad (4.9)$$

- Les phases d'affectation et de mise à jour sont répétées de façon itérative jusqu'à ce que l'erreur résiduelle soit inférieure à un seuil fixé.

4.4 Extraction de variables

La phase d'extraction de caractéristiques constitue généralement l'une des phases les plus importantes dans l'élaboration du traitement d'image. Il s'agit en effet de déterminer un espace numérique de description dans lequel les données images seront projetées et permettant une séparation optimale des classes.

La caractérisation ou la description est une étape dans laquelle on cherche à produire le même processus d'analyse et d'interprétation qui sera le plus proche que possible de « la description humaine ». La caractérisation a pour but d'extraire les caractéristiques qui décrivent au mieux et de façon quantitative et qualitative les objets présents dans l'image. Elle transforme les informations issues de la phase de segmentation en informations de telle sorte qu'elles soient décrites de façon analytique (Voir la Figure 4.8), les techniques de caractérisation sont variable et dépendent principalement de l'objectif visé et du type d'image que l'on veut analyser, nous trouvons des caractéristiques de forme, de couleur et de texture. Dans ce travail on s'intéresse aux caractéristiques de texture.

4.4.1 Les statistiques de premier ordre

Les méthodes statistiques du premier ordre donnent des informations générales, en caractérisant une image en fonction de la variation de l'intensité des pixels qui la compose [24].

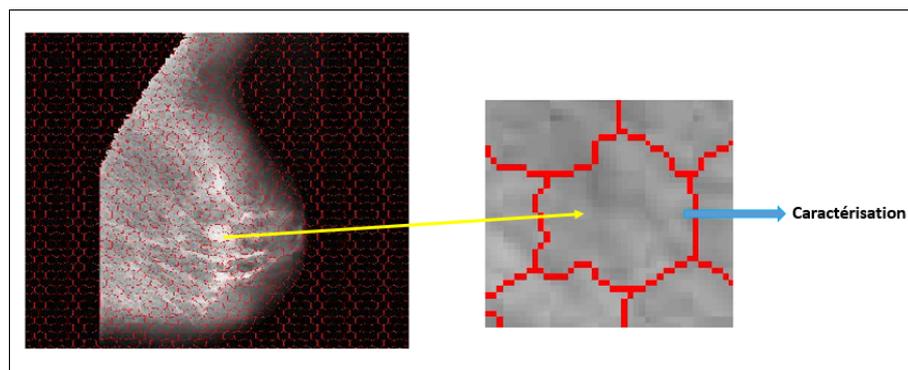


FIGURE 4.8 – Extraction de caractéristique.

La moyenne

La moyenne (m_1) est la valeur moyenne d'intensité dans un super-pixel, elle est calculée par la formule suivante :

$$m_1 = \frac{1}{M * N} \sum_{x=1}^M \sum_{y=1}^N I(x, y). \quad (4.10)$$

Avec, $[MN]$ est la taille de l'image et $I(x, y)$ est le pixel de l'image.

La médiane

La médiane (median) est la valeur qui divise exactement en deux la distribution la valeur d'intensité d'un super-pixel.

$$median = \left\{ \frac{i \left[\frac{M*N}{2} \right] + i \left[\frac{M*N}{2} + 1 \right]}{i \left[\frac{M*N}{2} + 1 \right]} \right\}. \quad (4.11)$$

Le mode

Le mode est la valeur des super-pixels de l'image dont la fréquence est la plus élevée.

$$mode = argmax_{i=0, \dots, 255} (H_I(i)). \quad (4.12)$$

La variance

Caractérise les variations des niveaux de gris (luminosité) par rapport à la valeur des niveaux de gris moyenne dans un super-pixels de l'image.

$$\sigma^2 = \frac{1}{M * N} \sum_{x=1}^M \sum_{y=1}^N (I(x, y) - m_1)^2. \quad (4.13)$$

Avec :

- M est la taille de l'image en x .
- N est la taille de l'image en y .
- $I(x,y)$ est le pixel « visité » à la position $\{x,y\}$.
- $m1$ est la moyenne.
- i : pixel.

4.4.2 Matrice cooccurrence

La matrice de cooccurrence a été introduite par Haralick en 1973 [25], elle est largement utilisée dans l'analyse de texture. Elle est très facile à mettre en œuvre et donne de bons résultats sur plusieurs types de texture. Dans la plupart des applications, les images utilisées sont représentées en niveaux de gris et la matrice de cooccurrence est connue sous le nom GLCM (Gray Level Co-occurrence Matrix).

Une matrice de cooccurrence est une matrice de taille $N \times N$, où N est le nombre de niveaux de gris présent dans une image. Pour un déplacement d (translation), un élément (i, j) de la matrice est défini par le nombre de pixels de l'image de niveau de gris j situés à d d'un pixel de niveau de gris i . La Figure 4.9 représente une transformation d'une image en GLCM.

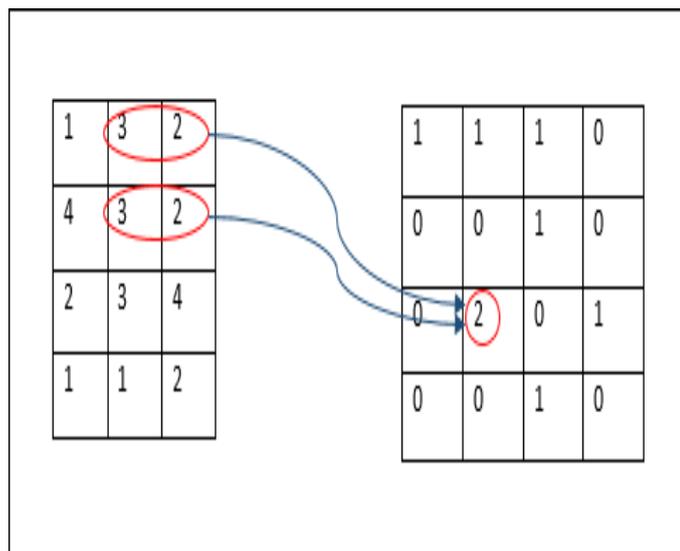


FIGURE 4.9 – Transformation d'une image en matrice de cooccurrence, (a) image et (b) la matrice de cooccurrence.

Les matrices de cooccurrence contiennent une masse très importante d'informations et sont donc difficilement manipulables, de ce fait, plusieurs indices ont été proposés par Haralick [25] qui correspondent à des caractères descriptifs de textures qui peuvent être calculés à partir de ces matrices comme : le contraste, la corrélation, l'énergie et l'homogénéité.

Dans ce qui suit, nous allons présenter les sept (7) caractéristiques que nous utiliserons durant le processus proposé :

Moyenne

Les valeurs d'intensité moyennes des pixels sont indiquées par cette caractéristique.

$$Moyenne = \sum_{i=0}^{N_G-1} \sum_{j=0}^{N_G-1} Mc(i, j) \quad (4.14)$$

avec : - $Mc(i,j)$: la matrice de cooccurrence de pixel (i,j) .

Écart-type

Il est beaucoup plus informatif, il indique la variabilité des niveaux de gris, c'est-à-dire, le changement du contraste.

$$\sigma_i^2 = \sum_{i=0}^{N_G-1} (i - \mu_i)^2 \sum_{j=0}^{N_G-1} Mc(i, j) \quad (4.15)$$

$$\sigma_j^2 = \sum_{j=0}^{N_G-1} (j - \mu_j)^2 \sum_{i=0}^{N_G-1} Mc(i, j) \quad (4.16)$$

Variance

Donne l'écart moyen des niveaux de gris par rapport à la moyenne c.à.d. il mesure l'ampleur des variations d'intensité par rapport à la moyenne.

$$variance = \sum_{i=0}^{N_G} \sum_{j=0}^{N_G} (i - moyenne)^2 Mc(i, j) \quad (4.17)$$

Energie

Mesure l'uniformité de la texture qui est la répétition des paires de pixels.

$$Energie = \sum_{i=0}^{N_G-1} \sum_{j=0}^{N_G-1} Mc(i, j)^2 \quad (4.18)$$

Contraste

Est une propriété intrinsèque d'une image qui quantifie la différence de luminosité entre les parties claires et les parties sombres d'une image.

$$Contraste = \sum_{i=0}^{N_G-1} \sum_{j=0}^{N_G-1} (i - j)^2 Mc(i, j) \quad (4.19)$$

Homogénéité

Plus on retrouve le même couple de pixels, plus cet indice est élevé, par exemple image uniforme, ou texture périodique dans le sens de la translation.

$$\sum_{ij} \frac{p(i, j)}{1 + (i - j)^2}. \quad (4.20)$$

Corrélation

La corrélation mesure la dépendance linéaire des niveaux de gris de l'image. (Décrit les corrélations entre les lignes et les colonnes de la matrice de cooccurrence).

$$correlation = \sum_{i=0}^{N_G-1} \sum_{j=0}^{N_G-1} \frac{(i - \mu_i)(j - \mu_j)}{\sigma_i \sigma_j} Mc(i, j) \quad (4.21)$$

avec :

- N_G : niveau de gris.
- $p(i,j)$: pixel.
- σ : Ecart type.
- Mc : la matrice de cooccurrence.

4.4.3 Filtre de Gabor

Le filtre de Gabor est un filtre fréquentiel qui connaît un certain succès, dû à leur similitude avec le fonctionnement de la perception visuelle humaine. En effet, il est maintenant admis que 2 cellules adjacentes du cortex visuel sont assez bien modélisées par une paire de filtres en quadrature. Un autre avantage des filtres de Gabor est qu'ils sont les seuls à atteindre la limite dans le principe d'incertitude d'Heisenberg [26].

:

Les filtres de GABOR ou filtres gaussiens constituent une classe particulière des filtres linéaires, ce sont des filtres orientés. Ces filtres ont une réponse impulsionnelle de la forme :

$$h(x, y) = g(x', y') e^{2\pi j(U_x + V_y)} \quad (4.22)$$

où :

$$(x', y') = (x \cos \phi + y \sin \phi, -x \sin \phi + y \cos \phi) \quad (4.23)$$

$$g(x', y') = \frac{1}{2\pi\sigma^2} * [exp(-(x'/\lambda)^2 + y'^2)/2\sigma^2)] \quad (4.24)$$

La réponse impulsionnelle $h(x, y)$ est donc une fonction complexe sinusoidale modulée par une gaussienne bidimensionnelle de rapport d'axes λ , de facteur de

dilatation σ et où ϕ est l'orientation de l'axe x' par rapport à l'axe x [27]. La transformée de FOURIER correspondante est

$$H(u, v) = \exp\left(-2\pi^2\sigma^2 \left[(u' - U')^2 \lambda^2 + (v' - V')^2\right]\right). \quad (4.25)$$

où :

$$(u', v') = (u\cos\phi + v\sin\phi - u\sin\phi + v\cos\phi). \quad (4.26)$$

Où (U', V') est obtenu par rotation du point (U, V) par le même angle ϕ .

4.4.4 Local Binary Pattern

Le Local Binary Pattern (LBP) ou "motif de texture binaire local" est un opérateur local d'analyse de texture. L'opérateur décrit la structure locale de la texture, par le biais de motifs circulaires. Des contraintes sont appliquées aux motifs afin de les rendre robustes et invariants. Ces motifs sont obtenus pour chaque pixel de l'image en souillant chaque pixel du voisinage circulaire par rapport au pixel central. Ce voisinage est paramétré par sa distance R au pixel central et le nombre P de pixels présents (population du voisinage). R est considéré comme un paramètre de résolution spatiale et P un paramètre de résolution angulaire. Il est donc possible de combiner plusieurs résolutions de l'opérateur en faisant varier à la fois la résolution spatiale et la résolution angulaire. Le LBP est capable ainsi de détecter les micro structures comme les coins et les bords [22].

$$LBP(x_c, y_c) = \sum_{n=0}^s 2^n s(i_n - i_c) \quad (4.27)$$

Avec, i_c intensité du centre, i_n intensité du voisinage et (x_c, y_c) les coordonnées du centre.

Les LBP sont des caractéristiques utilisées afin d'analyser les textures dans les images numériques. Plus précisément, on les utilise ici pour mesurer les similarités de texture entre les super-pixels.

Le vecteur de fonctionnalité LBP [25], dans sa forme la plus simple (Voir la Figure 4.10), il est créé de la manière suivante :

- Diviser la fenêtre à examiner en cellules (par exemple 16 x 16 pixels pour chaque cellule).
- Comparer chaque pixel d'une cellule à chacun de ses 8 voisins (en haut à gauche, en milieu à gauche, en bas à gauche, en haut à droite, etc.). Suivre les pixels le long d'un cercle, c'est-à-dire dans le sens des aiguilles d'une montre ou dans le sens inverse des aiguilles d'une montre.
- Lorsque la valeur du pixel central est supérieure à la valeur du voisin, écrire "0". Sinon, écrire "1". Cela donne un nombre binaire de 8 chiffres (qui est habituellement converti en décimal pour plus de commodité).

- Calculer l’histogramme, sur la cellule, de la fréquence de chaque "nombre" survenant (c’est-à-dire, chaque combinaison dont les pixels sont plus petits et qui sont plus grands que le centre). Cet histogramme peut être vu comme un vecteur de caractéristiques à 256 dimensions.
- Normaliser éventuellement l’histogramme.
- Concaténer les histogrammes (normalisés) de toutes les cellules. Cela donne un vecteur de caractéristiques pour toute la fenêtre.

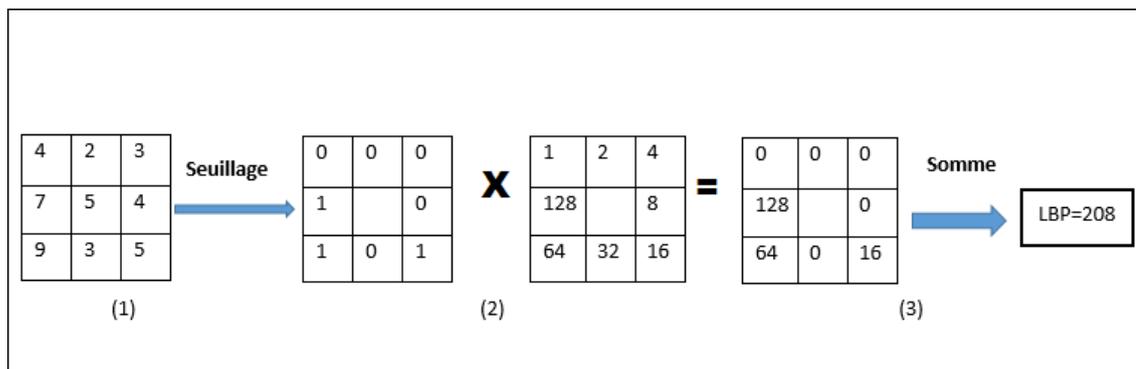


FIGURE 4.10 – Le Code LBP, (1) image en niveau de gris, (2) Multiplication de code binaire et le seuillage, (3) Représentation de LBP.

4.4.5 Moments de Hu

Les moments Hu [28] sont utilisés depuis longtemps pour calculer la position et le centre d’une distribution mais aussi sa variance. En vision par ordinateur, les moments permettent de calculer la position et l’orientation d’un objet. A partir des moments géométriques, HU a proposé un ensemble de sept moments invariants aux translations, rotations et changement d’échelle .

Une première utilisation des moments pour la reconnaissance de motifs géométriques a été proposée par HU [29]. Les moments offrent un cadre théorique puissant pour résoudre des problèmes rencontrés dans plusieurs applications d’imagerie y compris l’imagerie médicale, surtout lorsqu’il s’agit d’observer un objet rigide (comme les vertèbres) à différents niveaux d’échelle, rotations et translations dans une image 2D.

Les moments Hu sont un ensemble de 7 nombres calculés à l’aide des moments centraux invariants des transformations d’image. Les 6 premiers moments se sont avérés invariants en translation, en ampleur, en rotation et en réflexion. Alors que le signe du 7ème moment change pour la réflexion de l’image.

Les 7 moments sont calculés en utilisant les formules suivantes :

$$h_0 = \eta_{20} + \eta_{02}$$

$$h_1 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2.$$

$$h_2 = (\eta_{30} - 3\eta_{02})^2 + (3\eta_{21} - \eta_{03})^2.$$

$$h_3 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2.$$

$$h_4 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \dots\dots$$

$$h_5 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})].$$

$$h_6 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})[3(\eta_{21} + \eta_{12})^2].$$

(4.28)

avec η : Les moments centraux normalisés.

$$\mu_{pp} = \sum_{x=0}^M \sum_{y=0}^N g(x, y)(x - x_g)^p (y - y_g)^p$$

$$\eta_{pp} = \frac{\mu_{pp}}{\mu_{r00}}; y = \frac{p+p}{2} + 1.$$

4.5 Classification supervisée

Le processus de classification nécessite toujours une base d'apprentissage comme entrée. Créer une base d'apprentissage c'est avoir des individus, dont on connaît avec certitude la classe d'apprentissage.

L'apprentissage supervisée, qui comprend une série de méthodes capables d'apprendre à prédire des comportements ou à classer des données, à partir des images mammaires.

4.5.1 Arbre de décision

Un arbre de décision et comme son nom le suggère, est un outil d'aide à la décision qui permet de répartir une population d'individus en groupes homogènes selon des attributs discriminants en fonction d'un objectif fixés. Djebbar et al [17] ont utilisés le classifieur d'arbre de décision sur la base de données MIAS. Il permet d'émettre des prédictions à partir des données connues sur le problème par réduction, niveau par niveau, du domaine des solutions, d'autre façon c'est l'ensemble de règles de classification basant leur décision sur des tests associés aux attributs, organisés de manière arborescente.

L'arbre de décision [30] est une méthode qui a l'avantage d'être lisible pour les analystes et permet de déterminer les couples "attribut / valeur" discriminants à

partir d'un très grand nombre d'attributs et de valeurs.

- Un arbre est constitué de nœuds connectés entre eux par des branches.
- Une branche entre deux nœuds est orientée : l'un des nœuds de la connexion est dit nœud parent, et l'autre nœud enfant.
- Chaque nœud est connecté à un et un seul nœud parent, sauf le nœud racine qui n'a pas de parent.
- Chaque nœud peut être connecté à 0 ou n nœuds enfants.
- Les deux caractéristiques précédentes font qu'un arbre n'est pas un réseau ou graphe.
- Un nœud qui n'a pas de parents est appelé nœud racine ou racine.
- Un nœud qui n'a pas de nœuds enfants est appelé nœud feuille ou feuille.

La Figure 4.11 montre une illustration d'un arbre de décision.

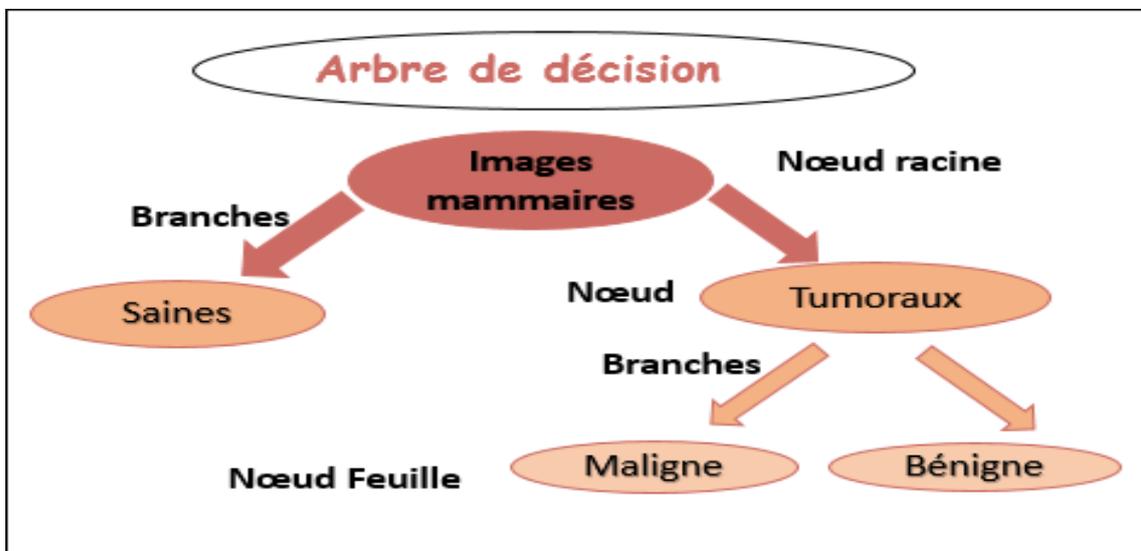


FIGURE 4.11 – Arbre de Décision

Chaque nœud interne d'un arbre de décision porte sur un attribut discriminant des éléments à classifier qui permet de répartir ces éléments de façon homogène entre les différents fils de ce nœud, les branches liant un nœud à ses fils représentent les valeurs discriminantes de l'attribut du nœud, et enfin, les feuilles d'un arbre de décision sont les prédictions des données à classifier (Algorithme 2.

Algorithme 2 Algorithme d'arbre de décision

- 1: On considère un nœud
 - 2: On sélectionne un attribut pour ce nœud
 - 3: On crée une branche pour chaque valeur de cet attribut
 - 4: Pour chaque branche, on regarde la pureté de la classe obtenue
 - 5: On décide si on termine la branche ou non
 - 6: Si on ne termine pas le processus est répété
-

4.5.2 Support Vector Machine (SVM)

Le SVM (Support Vector Machine) est une méthode pour estimer la fonction qui classe les données en deux classes. SVM est basé sur l'idée de minimiser l'erreur de généralisation lorsque le classifieur est appliqué à des échantillons de test ne correspondant à aucun échantillon d'apprentissage utilisé pour former le classifieur comme [6] en utilisant le classifieur SVM sur la base de données IN-BREAST.

Cette méthode cherche un hyperplan qui sépare les deux classes (cancéreuse et non cancéreuse), en garantissant que la marge entre le point de données le plus proche des deux classes soit maximale. L'intérêt de cette méthode est la sélection de vecteurs supports qui représentent les vecteurs discriminant grâce auxquels est déterminé l'hyperplan.

l'algorithme 3 se base principalement [31] sur 3 astuces pour obtenir de très bonnes performances tant en qualité de prédiction qu'en complexité de calcul.

Algorithm 3 Algorithme de SVM

- 1: On cherche l'hyperplan comme solution d'un problème d'optimisation sous-contrainte.
 - 2: Le passage a la recherche de surfaces séparatrices non linéaires est introduit en utilisant un noyau qui code une transformation non linéaire des données.
 - 3: Numériquement, toutes les équations s'obtiennent en fonction de certains produits scalaires utilisant le noyau et certains points de la base de données (ce sont les Support Vectors).
-

4.6 Conclusion

La classification super-pixellique est devenue une pratique très utilisée dans le domaine de la segmentation. Son principe se base sur la classification super-pixellique où chaque super-pixel est affecté à une classe de région. La segmentation super-pixellique a une capacité à délimiter les contours des régions incluses dans l'image, et cela a permis de traiter des super-pixels au lieu des pixels dans de nombreuses applications de vision par ordinateur.

Dans la procédure de traitement, on a fait appel à l'algorithme SLIC pour la génération des super-pixels, deux contraintes doivent être fixées par l'utilisateur, la compacité m pour contrôler la similarité du super-pixel, et le paramètre k pour simuler le nombre de super-pixel. Après cette étape, on passe à l'extraction des caractéristiques en se basant sur l'analyse de la texture de chaque super-pixel.

Une dernière étape consiste à réaliser la classification supervisée, en exploitant les caractéristiques extraites pour classer chaque super-pixel en super-pixel-fond, super-pixel-sein ou super-pixel-masse.

Chapitre 5

Résultat et discussion

Dans le chapitre précédent nous avons décrit la méthodologie que nous proposons pour réaliser la détection des masses mammaires à partir des images de mammographie, notre chaîne de traitement est illustrée dans Figure 5.1. Cette figure résume les techniques de notre processus, qui utilise le seuillage, l'égalisation adaptative d'histogramme et la morphologie mathématique pour la préparation et l'amélioration de l'image, puis nous avons réalisé une segmentation de l'image à base de super-pixels, par la suite nous avons appliqué une description de caractéristiques à base de LBP, moments Hu, filtre de Gabor, matrice de cooccurrence et des statistiques du premier ordre. Dans la dernière étape, nous avons procédé à une comparaison de deux méthodes de classification : machines à vecteurs de supports et arbre de décision. Ce chapitre représente les résultats obtenus avec la méthode proposée.

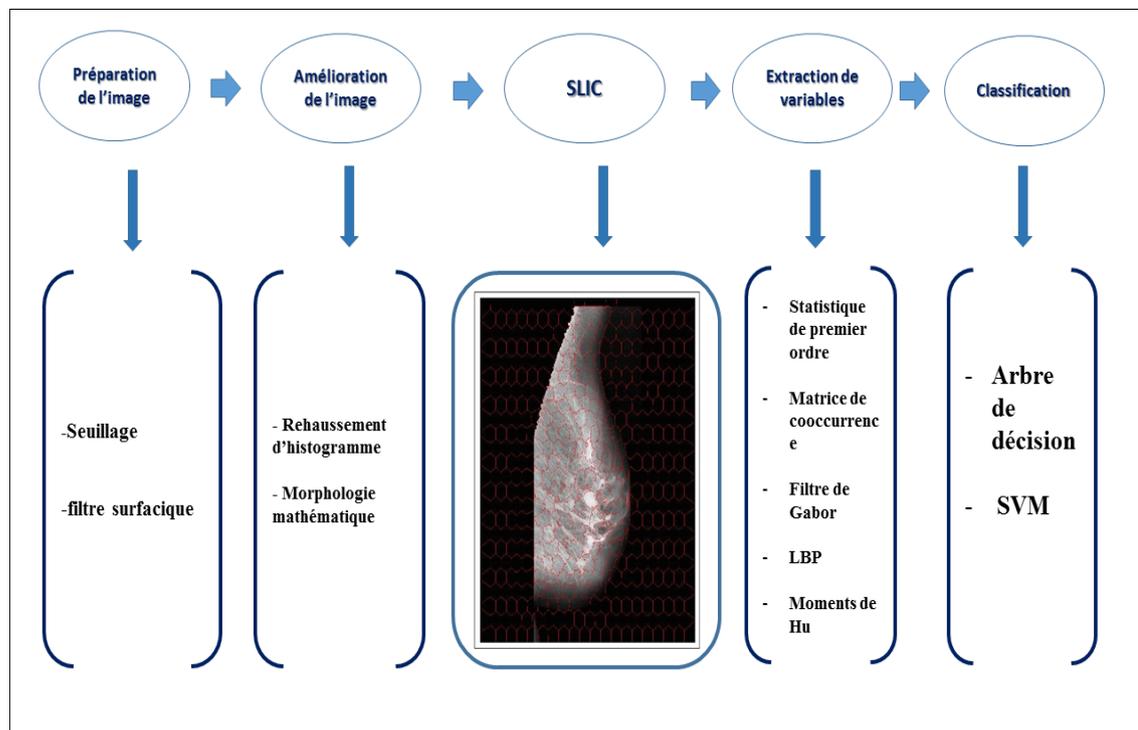


FIGURE 5.1 – Schéma résume notre méthode proposée .

5.1 Base de données

Dans ce travail, nous avons exploité la base de données MIAS [32], qui appartient à une organisation de groupe de recherche britannique intéressé par la compréhension des mammographies, cette organisation a généré une base de données de mammographies numériques qui contient 322 films numérisés. Le tableau 5.1 résume les classes de cette base de donnée.

Classe	Nombre d'image
Normal (Norm)	212
Calcification (CALC)	25
Distorsion architecturale (ARCH)	17
Autres masses mal définies (MISC)	14
Masses bien définies / circonscrites (CIRC)	24
Asymétrie (ASYM)	14
Masse spiculée (SPIC)	16
Total	322

TABLE 5.1 – Les classes de la base de données MIAS.

5.1.1 Description de la base de données

Le tableau 5.2 répertorie les films de la base de données MIAS et fournit les détails appropriés.

Numéro de référence	Exemple : mdb00, mdb322 ...
Classe d'anomalie	- Gras - Gras - glandulaire - Dense - glandulaire
Gravité de l'anomalie	- CALC : Calcification - CIRC : Masses bien définies / circonscrites - SPIC : Masse spiculée - MISC : Autres masses mal définies - ARCH : Distorsion architecturale - ASYM : Asymétrie - NORM : Normal
Coordonnées du centre d'anomalie	x,y, exemple : x=450, y=900
Rayon approximatif (en pixels) d'un cercle contenant l'anomalie	Exemple : r= 50
Type	- B : bénigne - M : maligne

TABLE 5.2 – Description de la base de données.

5.2 Paramétrages

Chaque technique utilisée dans notre travail nécessite un paramétrage, le tableau 5.3 résume les paramètres utilisés qui ont été fixés selon la littérature ou après plusieurs tests de notre part.

Préparation de l'image	Seuillage : Seuil = [0.55 : 0.89]
Amélioration de l'image	ES : (line, 10 ,45), (DISK ,3)
Segmentation	K = 2000 et m = 15
Extraction de variables	LBP : 8 voisins, Filtre de Gabor : 5 masques gaussiens

TABLE 5.3 – Paramétrage.

5.3 Résultats et discussions

5.3.1 Préparation de l'image

Les images brutes présentent beaucoup d'artéfacts qu'il faut les enlever, comme les étiquettes, le rai de numérisation et le muscle pectorale, la Figure 5.2 représente le résultat de cette étape. Nous avons procédé cette préparation comme suit :

- Elimination des étiquettes par opération de masquage.
- Elimination de muscle pectoral par une opération de seuillage (un seuil de 0.68), suivi d'un filtrage surfacique.

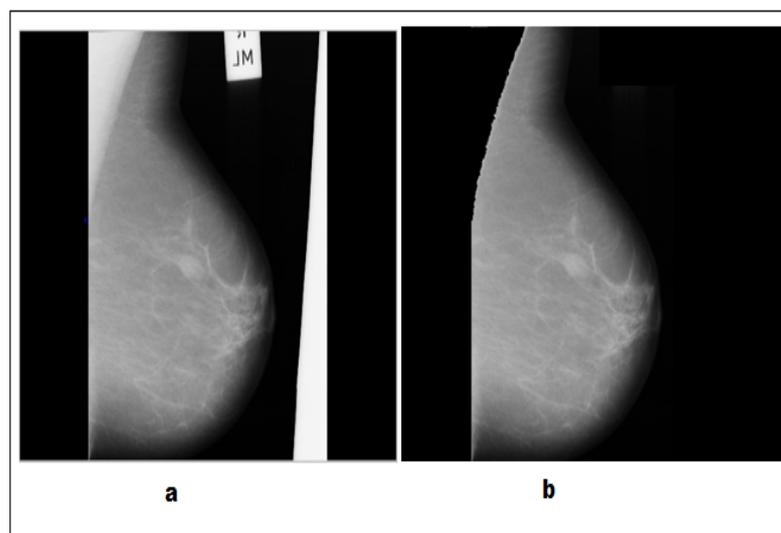


FIGURE 5.2 – (a) Image brute et (b) Image Résultante.

5.3.2 Amélioration de l'image

Dans cette étape nous avons appliqué 5 méthodes d'amélioration d'image, qui sont : l'ouverture, la fermeture, WTH, BTH et égalisation d'histogramme adaptatif.

Pour les opérations morphologiques, nous avons réalisé plusieurs tests avec différents éléments structurants de taille différente. Le premier élément structurant est sous forme d'un disque avec des rayons différents (SE = disque, rayon = 1,2 et3). Le deuxième élément structurant est sous forme d'une ligne dont nous avons changé à chaque fois sa taille et son orientation (SE = line (10,45), (40,120), (120,20)).

Dans cette expérimentation, nous avons testé le résultat obtenu par le PSNR (Peak Signal to Noise Ratio) qui est un test comparatif permettant de mesurer la dégradation entre l'image originale et l'image résultante. Le calcul de PSNR nécessite le calcul de l'erreur quadratique moyenne EQM (Evaluation de qualité : entre une image source I (de taille N*M) et une image résultante I' (de même taille)).

$$PSNR_{db} = 10 \log_{10} \left(\frac{l^2}{EQM} \right). \quad (5.1)$$

Avec : l le nombre d'échelons de luminance (255 pour des images codées sur 8 bits) .

Une valeur de PSNR inférieure à $30dB$ traduit généralement une image qui présente des dégradations perceptibles.

La méthode qui présente le plus grand PSNR est généralement la meilleure méthode qui résulte une bonne amélioration d'image. D'après le tableau 5.3.2, l'égalisation adaptative d'histogramme est la méthode qui présente le plus grand PSNR qui égale à $41.46dB$, la deuxième classe a pour l'opération morphologique d'ouverture avec un PSNR de $40,16dB$, et pour la troisième classe BTH avec un PSNR de $10,39dB$. La Figure 5.3 présente le résultat qualitatif de chaque technique d'amélioration.

IMAGES	SE='DISK'=3				SE=line='10,45'				BHLE
	O	F	WTH	BTH	O	F	WTH	BTH	
Image 01	37.01	10.38	10.39	10.42	40,09	10,38	10,39	10,41	40,53
Image 02	39.05	11.08	11.11	11.14	37,71	11,08	11,11	11,15	37,91
Image 03	38.98	10.57	10.59	10.62	39,8	10,57	10,6	10,62	40,92
Image 04	36.63	10.52	10.54	10.57	34,66	10,52	10,54	10,58	41
Image 05	35.82	8.59	8.62	8.65	36,69	8,59	8,61	8,64	38,92
Image 06	37.21	10.05	10.08	10.11	37,96	10,05	10,07	10,1	43,47
Image 07	40.33	8.26	8.27	8.28	42,84	8,26	8,26	8,28	39,3
Image 08	38.48	11.75	11.79	11.82	35,26	11,03	11,08	11,13	38,51
Image 09	38.56	14.59	14.66	14.70	37,44	11,75	11,8	11,85	38,43
Image 10	38.55	11.59	11.64	11.67	38,17	7,43	7,44	7,46	39,39
Image 11	42.53	11.78	11.80	11.82	38,44	11,59	11,64	11,67	43,87
Image 12	38.90	12.72	12.75	12.79	40,53	11,78	11,8	11,83	42,02
Image 13	43.02	8.72	8.74	8.76	41,2	8,72	8,75	8,77	42,98
Image 14	38.78	7.43	7.45	7.46	40,47	7,51	7,54	7,55	43,14
Image 15	42.72	7.51	7.53	7.54	40,91	9,69	9,72	9,73	44,79
Image 16	44.51	9.69	9.71	9.72	38,49	10,9	10,93	10,96	37,74
Image 17	41.05	7.98	7.99	8.01	44,45	13,62	13,64	13,66	48,75
Image 18	46.83	13.62	13.64	13.66	42,54	9,57	9,58	9,59	44,53
Moyenne	40.16	10.34	10.36	10.39	39,31	10,17	10,1997	10,22	41,46

TABLE 5.4 – Synthèse de PSNR.

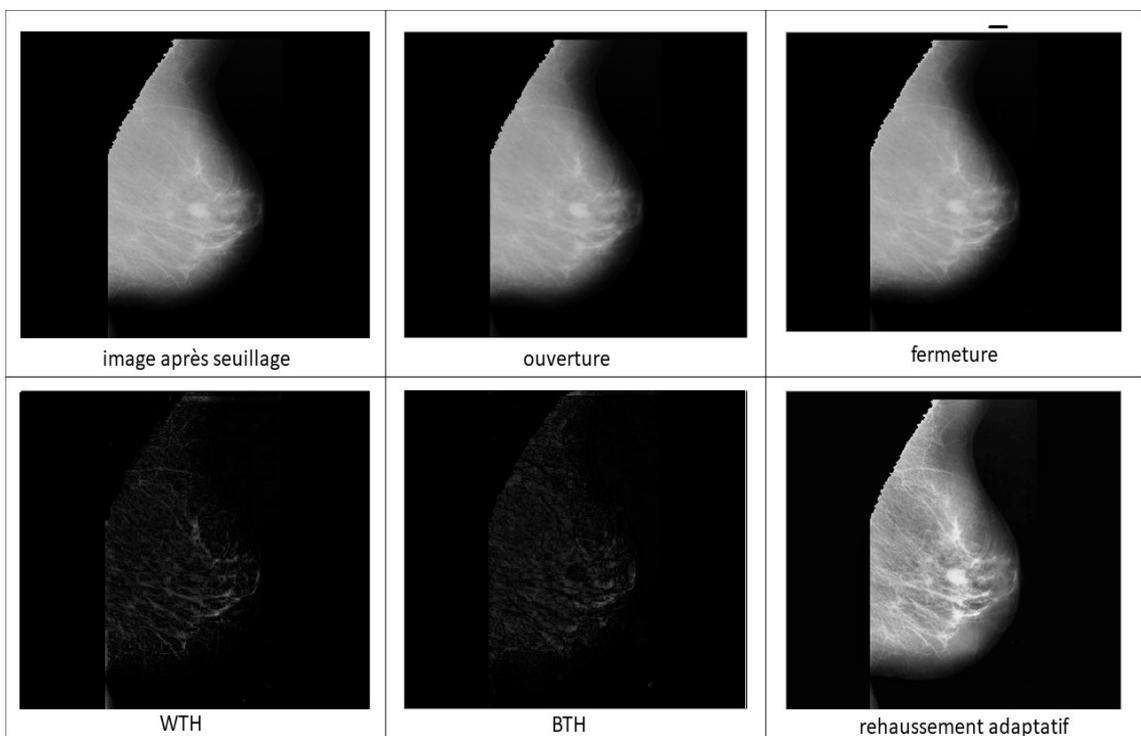


FIGURE 5.3 – Les résultats de chaque technique d'amélioration.

5.3.3 SLIC

Nous avons choisi la méthode de super-pixel qui repose sur l'algorithme SLIC, dont on a changé à chaque fois le paramètre k ($k = 500, 1500, 2000$) et $m = 15$.

Selon les résultats obtenus, nous avons choisi $k=2000$, parce qu'il donne une meilleure perception de super-pixels. La figure 5.4 peut nous confirmer ce choix.

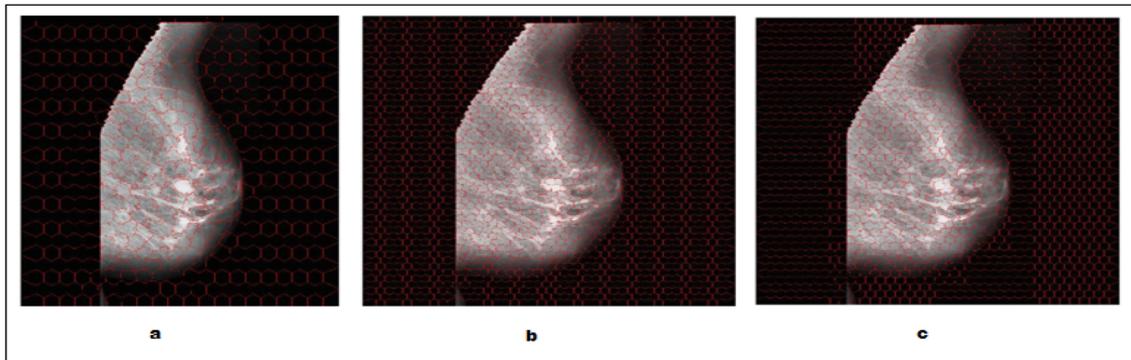


FIGURE 5.4 – Résultat de l'algorithme SLIC, (a) $k=500$ super-pixels, (b) $k=1500$ super-pixels et (c) $k=2000$ super-pixels.

5.3.4 Détection automatique des masses mammaires

En premier lieu, nous devons appliquer une caractérisation de texture en calculant les Statistique de premier ordre, MC, Filtre de Gabor, LBP et Moment Hu, afin de réaliser notre base d'informations qui nous permette d'attaquer la phase de classification.

En deuxième lieu, nous avons réalisé une classification de chaque super-pixel de l'image, nous avons testé deux classifieurs : l'arbre de décision (classification tree) et la classification par SVM.

Afin de comparer le résultat de la classification et de la segmentation, nous avons calculé le paramètre $F - score$ et le taux de classification (Accuracy ACC).

$$F - score = 2 * (Precision * Rappel) / (Precision + Rappel). \quad (5.2)$$

$$Acc = (VP + VN) / (VP + VN + FP + FN). \quad (5.3)$$

Avec :

Précision = $VP / (VP + FP)$

Rappel = $VP / (VP + FN)$.

Le tableau 5.5 présente les résultats obtenus de la classification. Nous remarquons que les résultats de taux de classification sont très proches pour les deux méthodes de classification, $ACC = 0.9979$ pour le classifieur d'arbre de décision, et $ACC = 0.9977$ pour le classifieur SVM. En revanche, nous remarquons qu'il y'a

une différence de performance au niveau de $F - score$, $F - score = 0.4810$ pour le classifieur d'arbre de décision, et $F - score = 0.4413$ pour le classifieur SVM

Les résultats obtenus nous ont incités à choisir la méthode d'arbre de décision comme le meilleur moyen de classification.

images	ACC SVM	ACC Arb	$F - score$ SVM	$F - score$ Arb
Image 01	0.9984	0.9984	0.6071	0.7120
Image 02	0.9987	0.9991	0.6347	0.7979
Image 03	0.9979	0.9976	0.7237	0.5299
Image 04	0.9979	0.9976	0.4950	0.5299
Image 05	0.9985	0.9977	0.6953	0.5389
Image 06	0.9995	1	0.8778	1
Image 07	0.9975	0.9970	0.6093	0.4345
Image 08	0.9982	0.9961	0.6843	0.0202
Image 09	0.9954	0.9959	0.2288	0.3636
Image 10	0.9990	0.9996	0.5604	0.8624
Image 11	0.9984	0.9938	NaN	NaN
Image 12	0.9968	0.9977	NaN	NaN
Image 13	0.9968	0.9968	0.0115	0.0115
Image 14	0.9991	0.9989	0.7059	0.6249
Image 15	0.9987	0.9984	0.3975	0.6151
Image 16	0.9962	0.9984	0.4655	0.6151
Moyenne	0.9979	0.9977	0.4810	0.4413

TABLE 5.5 – Accuracy et le F-score de SVM et Arbre de décision.

5.3.5 Test d'efficacité de la méthode proposée

Pour assurer que notre programme fonctionne correctement, nous avons réalisé un test sur des images qui contient des masses mammaires, et notre programme a pu classer ses images comme des images cancéreuses en segmentant les bords de la masse mammaire.

Comme nous avons réalisé un deuxième test sur des images qui ne contient aucune masse mammaire, et notre programme a pu les classer comme des images saines sans aucune segmentation car elles ne contiennent pas des masses mammaires. Les résultats de ce test sont illustrés dans la Figure 5.5.

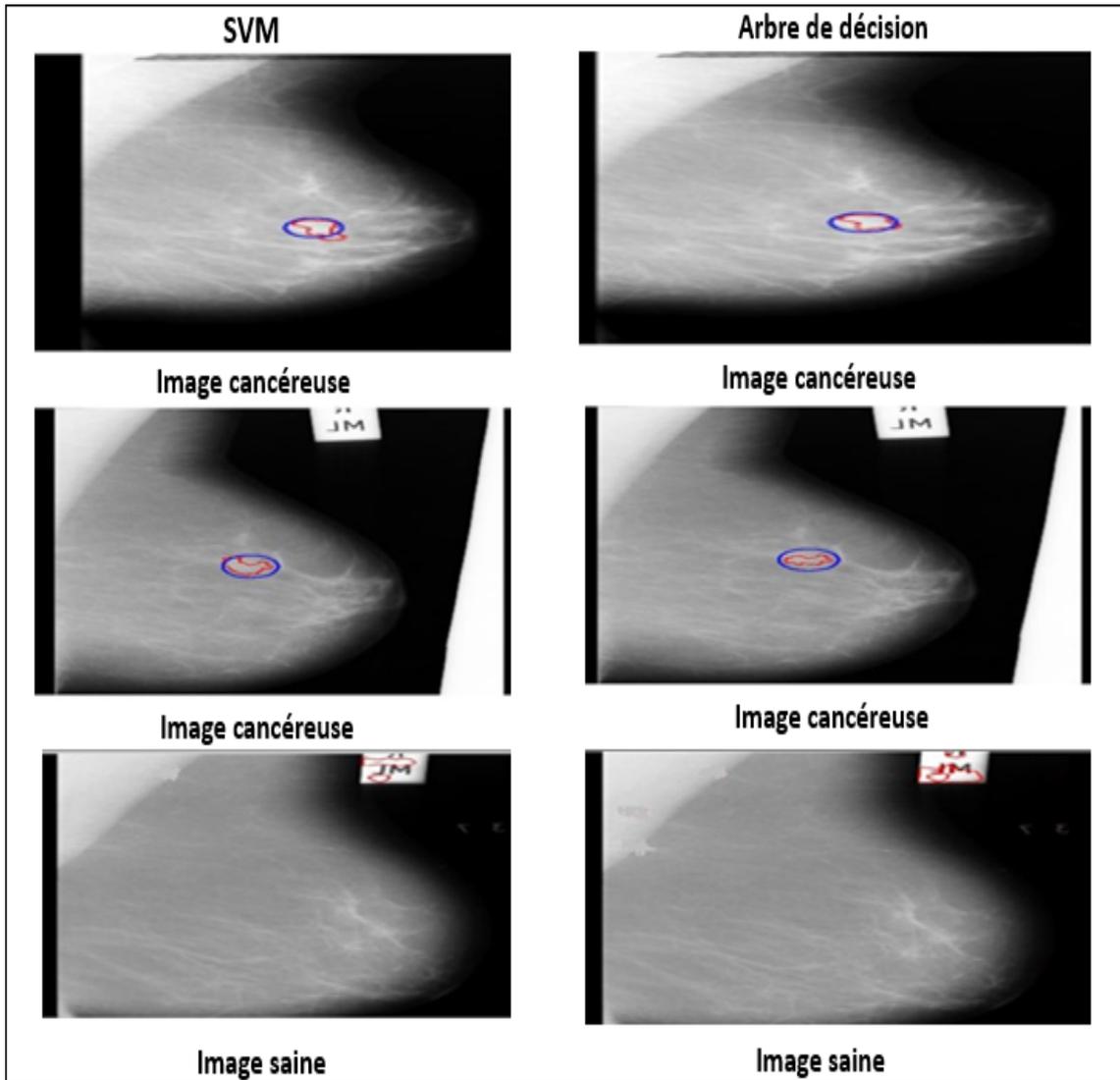


FIGURE 5.5 – Résultat comparatif entre SVM et l'arbre de décision

5.3.6 Présentation de l'application "Inter"

Pour concrétiser notre travail, nous avons développé une application sous le nom « Inter » qui synthétise notre approche. La fenêtre de notre application est composée d'un Panel d'option, deux axes d'affichage et une zone de texte.

Notre application est réalisée sur un PC de type hp i3-5005u, 4Go de Ram, sous windows 10. Les langages utilisés sont : Matlab 2014.

Le Panel d'option contient cinq boutons :

1. Le bouton "Lire image" permet la lecture et l'affichage de l'image. L'image sera affichée sur l'axe 1 (image brute).
2. Le bouton "C/S par SVM" permet de faire une classification par la méthode

SVM, le résultat de segmentation sera affiché sur l'axe deux, et les bornes de la masse mammaire seront indiquées en contour rouge.

3. Le bouton "C/S par Arbre de Décision" permet de faire une classification par le classifieur d'arbre de décision, le résultat de segmentation sera affiché sur l'axe deux, et les bornes de la masse mammaire seront indiquées en contour rouge.
4. Le bouton "save" permet d'enregistrer l'image résultante sous l'extension JPEG.
5. Le bouton "Quit" permet de quitter l'application.

L'Edit text permet d'afficher le commentaire de l'image si elle est saine ou cancéreuse. La figure 5.6 présente une simulation d'une classification par SVM.

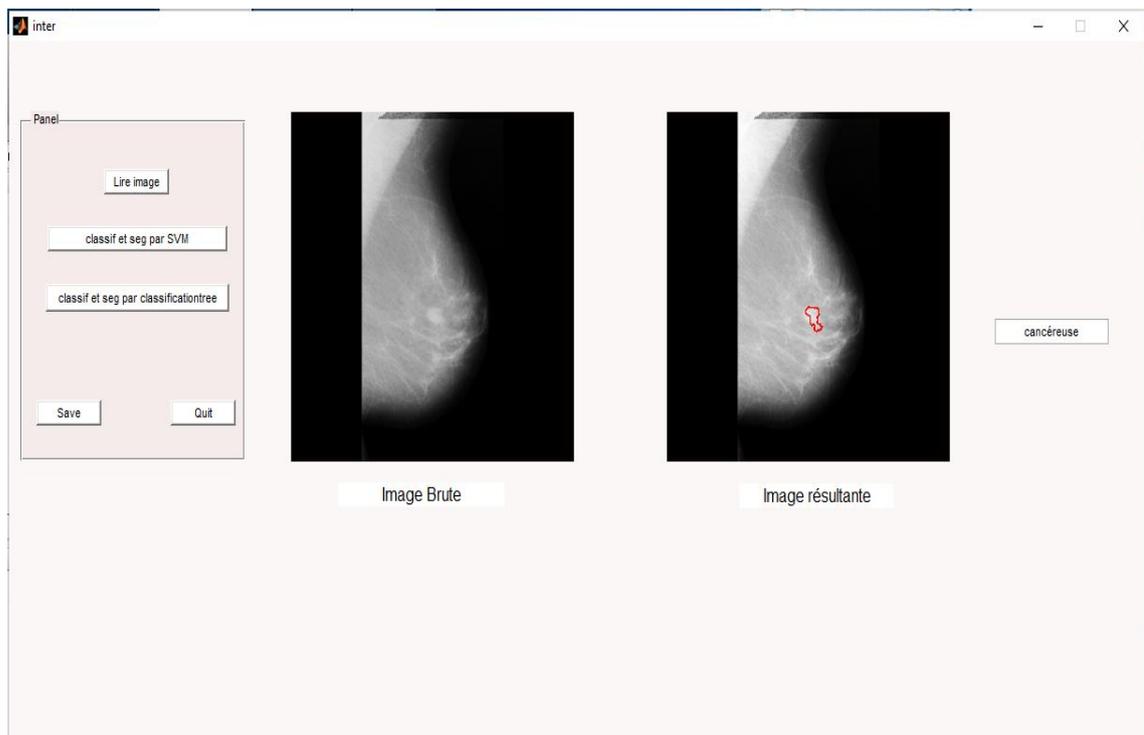


FIGURE 5.6 – Classification par SVM.

5.4 Conclusion

Dans ce chapitre nous avons décrit notre système de classification en présentant les différentes méthodes utilisées pour la classification des masses mammaires.

D'après les résultats, l'égalisation d'histogramme en prétraitement, la segmentation par SLIC, et l'arbre de décision, sont les méthodes les plus adéquates pour

notre travail.

Chapitre 6

Conclusion

Dans ce travail, on s'est intéressé à l'étude des systèmes automatiques de diagnostic assistés par ordinateur en vue de détection précoce du cancer du sein, notamment à la description et à la classification des images mammaires. En effet, les statistiques confirment que le cancer du sein représente une menace prépondérante pour la vie de la femme. Cependant, une telle menace n'est gérée que par la prise en charge rapide de la maladie pour maximiser les chances de survie. D'où, le recours au moyen informatique automatisant la procédure de diagnostic appelé : le Diagnostic Assisté par Ordinateur (DAO). Cette procédure permet de révéler des détails difficiles à percevoir à l'œil nu, tout en évitant la création d'artefacts, faussement informatifs.

Ce qui justifier notre choix de seuillage puisqu'il permet de prendre efficacement, des notions de taille, de forme, de contraste et conserve l'information sur le contour, sont des notions importantes dans le diagnostic des masses mammaires.

Aussi nous avons utilisé l'algorithme d'égalisation d'histogramme qui nous a permis d'améliorer la segmentation des masses mammaires.

La méthode que nous avons choisi est la segmentation super-pixellique par l'algorithme SLIC, qui nécessite le choix de deux paramètres : la quantité de super-pixels k et m qui correspond au poids attribué à la distance spatiale. Le bon choix de ces paramètres permet la détection des régions d'intérêt avec des résultats très satisfaisants, en termes de qualité de segmentation et en de temps d'exécution.

Dans une deuxième phase, nous nous sommes intéressées aux différentes techniques d'extraction de variables afin de classifier les masses mammaires selon la texture de chaque super-pixel. L'application de ces fonctions a fourni des résultats très satisfaisants, pour caractériser les images mammaires (cancéreuses et saines).

Dans la dernière phase de notre travail, nous avons décrit en détails les étapes de classification par la méthode de SVM et de l'arbre de décision, ainsi que les résultats obtenus.

D'après les résultats obtenus, nous pouvons conclure que la méthode d'arbre de décision est qualifiée comme meilleure et satisfaisante grâce à la classification précise de la masse mammaire.

A la lumière de ce qui a été fait, ce travail est amené à être prolongé dans plusieurs directions, des idées futures se dessinent :

Appliquer la méthode sur d'autre type d'anomalies mammaires (microclassification , macroclassification) et avec d'autre base de données. En se basant sur des caractérisation de la forme et de la transformée de fourrier au niveau d'extraction de variables, avec une grande base de données dans l'apprentissage pour augmenter les performances de la classification.

Améliorer l'application au niveau de la classification pour préciser si la masse mammaire maligne ou bien bénigne .Et comparer les résultats avec d'autres méthodes peut amener à sélectionner le meilleur modèle approprié à la détection.

Bibliographie

- [1] H.TERKI et K BENYELES, "analyse des images mammothographiques en vue de la detection et la caractérisation des microcalcifications mammaires," M.S. thesis, 2013.06.23.
- [2] Journal de Liberté Algérie, "http://www.ecoles.cfwb.be.htm," Consulté le : 25/06/2019.
- [3] C.DIAZ-HUERTA, F.RIVERON EDGADRO, and M.ZETINA, "Quantitative analysis of morphological techniques for automatic classification of microcalcifications in digitized mammograms," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7361–7369, 2014.
- [4] Institut national de cancer, "Octobre 2013.https://www.e-cancer.fr/patients-et-proches/les-cancers/cancer-du-sein/references," Consulté le : 12/07/2018.
- [5] DR.K.FARHANI, "cours de mammothographies," 2017/2018.
- [6] N. El ATLASand A. BYBI and H. DRISSI, "Features fusion for characterizing inbreast-database masses," in *2016 International Conference on Electrical and Information Technologies (ICEIT)*, May 2016, pp. 374–379.
- [7] E Abd El-Samie, "An algorithm for pre-processing and segmentation of mammogram images," in *Computer Engineering & Systems (ICCES), 2016 11th International Conference on*. IEEE, 2016, pp. 187–190.
- [8] L.XIAOMING and T.JINSHAN, "Mass classification in mammograms using selected geometry and texture features, and a new svm-based feature selection method," *IEEE Systems Journal*, vol. 8, no. 3, pp. 910–920, 2014.
- [9] H. BYUNG-WOO and S.BONG-SOO, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 129–139, 2010.
- [10] Kai Hu, Xieping Gao, and Fei Li, "Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 2, pp. 462–472, 2011.
- [11] F.CHOKRI and F. MEROUANI, "Mammographic mass classification according to bi-rads lexicon," *IET Computer Vision*, vol. 11, no. 3, pp. 189–198, 2016.
- [12] S.SHAFIULLAH and C.KWANG NAM, "Robust active contours for mammogram image segmentation," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2149–2153.

- [13] S.PENG, Z. JING, R. ANDRILAK, and W. HUI, "A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms," *Computers in biology and medicine*, vol. 96, pp. 178–188, 2018.
- [14] O.MARTINS, J.GERADLO BRAZ, S. ARISTÓFANES CORRÊA, P. ANSELMO CARDOSO, and G. MARCELO, "Detection of masses in digital mammograms using k-means and support vector machine," *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 8, no. 2, pp. 39–50, 2009.
- [15] R. JINCHANG, "Ann vs. svm : Which one performs better in classification of mccs in mammogram imaging," *Knowledge-Based Systems*, vol. 26, pp. 144–153, 2012.
- [16] H.RAHLI and N.BENAMRANE, "Interpretation des images de mammographie par l'algorithme search harmony.," *Annals. Computer Science Series*, vol. 10, no. 2, 2012.
- [17] A.DJEBBAR, H. DJELLALI, and H. MEROUANI, "Classification des tumeurs du sein par un classifieur bayésien naïf et les arbres de décision," .
- [18] C.JINGHUI, M.HANG, L.LI, and L.WEI, "A novel computer aided breast mass detection scheme based on morphological enhancement and slic superpixel segmentation," *Medical physics*, vol. 42, no. 7, pp. 3859–3869, 2015.
- [19] GUERRODJI, "Segmentation des clichés mammographiques en vue de la dtecton des foyers de micro-calcifications mammaires," M.S. thesis, Université Mouloud Mammeri, Tizi - Ouzou.
- [20] M.CHABANE et S.HADDACHE, "segmentation des images mammographies," M.S. thesis, universitée de bijaia abderrahmane mira.
- [21] W. HONGYU, F.JUN, B. QIRONG, L. FEIHONG, Z. MIN, R. YU, and L. YI, "Breast mass detection in digital mammogram based on gestalt psychology," *Journal of healthcare engineering*, vol. 2018, 2018.
- [22] C.Desir, "Classification automatique d'images, application ál'imagerie du poumon profond :<https://tel.archives-ouvertes.fr/tel-00879356/document>," Consulté le : 15/05/2019.
- [23] N.ZEROUAL, "Analyse des images multi-spectrales de la peau pour l'aide au depistage du melanome," M.S. thesis, université Abou Bakr Belkaid de telemcen, 24 juillet 2018.
- [24] D.ANDERSON, "Analyses et seynthese de textures.2007-2008.<http://benjlaiel.hebergratuit.net/texture-rapport.pdf>," Consulté le : 01/06/2019.
- [25] A.MEKHMOUCH, "Segmentation d'image par classificasson de pixel : application aux images irm," M.S. thesis, 2017.
- [26] Futurasciences "<https://www.futura-sciences.com/sciences/definitions/physique-inegalites-heisenberg-4753/>," consult le : 20/03/2019.
- [27] Newtechnology "www.telemcen.ulg.ac.be/teaching/notes/totali/elen016/node46.html," consult le : 01/04/2019.

-
- [28] R.BOUMAZA et H. BOUDIAF, "Reconnaissance d'empreintes digitales en utilisant les moment hu," M.S. thesis, Universit l'arbi en mhidi oum el boua-ghi.
- [29] N.ACHOUR et T.DROUICHE, "Segmentation d'image par classification de pixel : application aux images irm," M.S. thesis, Université A.Mira de bejaia, 2018.06.26.
- [30] K.BOUDLAL, "Realisation d'une application pour la detction de type de la masse mammaire," M.S. thesis, Universié aboubakr belkaid telemcen, 2016.05.25.
- [31] S.Gadat, "Algorithmes de support vector machines.http ://www.math.univ-toulouse.fr / gadat/ens/m2sid/12-m2-svm.pdf," Consulté le : 20/04/2018.
- [32] A.CLARK, "The mini-mias database of mammograms.11 dmbre 2012 ,http ://peipa.essex.ac.uk/info/mias.html," Consulté le : 25/09/2018.