

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABOU BEKR BELKAID
FACULTÉ DE TECHNOLOGIE
DÉPARTEMENT DE GÉNIE BIOMÉDICAL

MÉMOIRE DE FIN D'ÉTUDES

pour obtenir le grade de
MASTER EN GÉNIE BIOMÉDICAL
Spécialité : **Informatique Biomédicale**
présenté et soutenu publiquement
par

Mr. Abderrezak MEZOUAGHI

et

Mr. Rachid NIATI

le 24 Décembre 2020

Titre:

Classification des images de mammographie

Jury

Présidente du jury. Dr. MEKKIOUI Nacira,
Examinatrice. Dr. SETTOUTI Nesma,

MCB UABB Tlemcen
MCA UABB Tlemcen

Directeur de mémoire. Dr. BECHAR Mohammed El Amine, MCB UABB Tlemcen

*Je dédie ce modeste travail et ma profonde gratitude
A tous celui qui a sacrifié pour m'offrir les conditions propices à ma réussite :*

***Mes parents, à ma chère mère et à l'âme de mon chère père Allah Y'rahmou, à qui je
dois la réussite, pour l'éducation qu'ils m'ont prodigué; avec tous les moyens et au prix
de toutes les sacrifices qu'ils ont consentis à mon égard, pour le sens du devoir qu'ils
m'ont enseigné depuis mon enfance,***

*A mon **cher frère**, mon bras droit qui me soutient tout le temps à tout prix,*

*A mon binôme **Rachid NIATI**, et à toute sa famille pour leur soutien et leur compassion,
Et à **Mes amis**, qui ont été ma seconde famille; Amine BRIKI, Ahmed ZAZIA,
Oussama BARROUCHI, Walid BOUZID et plein d'autres, pour leur chaleureuse
compagnie qui m'a aisé ce long trajet, avec lesquels j'ai vécu les plus beaux souvenirs, et
sur eux je sais que je peux toujours compter,*

*Mais avant et après tout, on dis **Alhamdulillah..***

Remerciements

*Nous voudrions dans un premier temps remercier, notre directeur de mémoire **Mr. BECHAR Mohammed El Amine** pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.*

Nous adressons nos sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé nos réflexions et ont accepté de nos rencontrer et de répondre à nos questions durant mes recherches.

*Nos vifs remerciements vont également aux membres du jury **Mme. MEKKIOUI Nawel.** et **Mme. SETTOUTI Nesma.** d'avoir accepté de corriger et d'évaluer notre travail.*

Résumé

La mammographie devient un examen de référence pour la détection du cancer du sein, même à un stade précoce, mais la tâche n'est toujours évidente due à plusieurs difficultés comme la présence des artéfacts ou bien la densité du glande mammaire elle-même. Vue ces challenges, une assistance d'un système informatique est toujours favorable.

A travers ce projet, nous voulons créer un système de classification qui aide effectivement au diagnostic du cancer du sein.

Notre approche consiste d'abord à appliquer des techniques de pré-traitement et de segmentation aux images de la base de données INbreast pour faciliter la tâche de classification. Ensuite, le résultat de la segmentation est suivi par une phase de caractérisation pour extraire des descripteurs d'image. A la fin, un classifieur supervisé est mise en place pour classer les masses mammaires en malignes ou bénignes tout en basant sur ces descripteurs. Les résultats obtenus peuvent être considérés comme prometteurs.

Mots-clés

Classification, supervisé, mammographie, cancer du sein, INbreast, segmentation, caractérisation, descripteurs, masses.

Abstract

Mammography becomes a benchmark examination for detecting breast cancer even at an early stage, but the task is not always obvious due to several difficulties such as the presence of artefacts or the density of the breast itself. In view of these challenges, support from a computer system is always advantageous.

With this project, we aim to create a classification system that efficiently helps in the diagnosis of breast cancer. Our approach is to first apply pre-processing and segmentation techniques on the images from the INbreast database to facilitate the task of classification. Then, the result of the segmentation is followed with a characterization phase to extract image descriptors. At the end, a supervised classifier based on these descriptors is set up to classify the breast masses into malignant or benign. The obtained results can be considered as promising.

Keywords

Classification, supervised, mammography, breast cancer, INbreast, segmentation, characterization, descriptors, masses.

Table des matières

Remerciements	i
Résumé	ii
Abstract	iii
Table des matières	iv
Table des figures	vi
Liste des tableaux	viii
Acronymes	ix
Introduction	1
1 Chapitre 01 : Informations générales	3
1 Introduction	3
2 L'anatomie du sein	3
3 Qu'est-ce que le cancer du sein?	4
4 Les différents types de tumeurs	5
4.1 Tumeurs bénignes	5
4.2 Tumeurs prémalignes	6
4.3 Tumeurs malignes (Cancéreuses)	6
4.4 Tumeurs phylodes	6
5 Définition de la mammographie	7
6 Les masses	8
6.1 La forme	8
6.2 Le contour	9
6.3 La densité	10
7 Qu'est-ce qu'un SAD?	11
7.1 Rôle d'un SAD	11
7.2 Principe d'un SAD	11
7.3 Différentes méthodes de l'aide au diagnostic	12
8 Conclusion	12
2 Chapitre 02 : L'état de L'art	13
1 Segmentation des masses mammaires	14
2 Classification des masses mammaires	17
3 Conclusion	18
3 Chapitre 03 : Partie théorique	19
1 Introduction	19
2 Prétraitement : Égalisation de l'histogramme	20
2.1 Histogramme d'une image	20

2.2	Égalisation de l'histogramme	20
3	Segmentation : Seuillage binaire	22
3.1	Segmentation	22
3.2	Binarisation-seuillage	22
4	Extraction de caractéristiques : GLCM	23
4.1	Caractéristiques d'une image	23
4.2	Extraction des caractéristiques	23
4.3	La matrice de co-occurrence de niveau de gris (GLCM)	24
4.3.1	Texture	24
4.3.2	GLCM	25
5	Classification : Par les forêts aléatoires (RF)	27
5.1	Classification	27
5.1.1	Classification supervisée	27
5.2	Forêts aléatoires (ou Random Forests)	27
6	Conclusion	29
4	Chapitre 04 : Partie expérimentale	30
1	Introduction	30
2	La base de données : INbreast	30
2.1	Nos données : Protocole de test	31
3	L'environnement de développement : Une brève description	31
3.1	Description du Hardware	31
3.2	Description du Software	32
3.2.1	Python	32
3.2.2	Anaconda : Jupyter notebook	32
3.2.3	Bibliothèques utilisées	33
4	Implémentation et résultats	34
4.1	Prétraitements	34
4.2	Segmentation	35
4.3	Caractérisation et apprentissage	36
4.4	Discussion des résultats	37
5	Conclusion	38
	Conclusion et perspectives	39
	Bibliographie	40

Table des figures

1.1	Une illustration présente l'anatomie du sein [6].	4
1.2	Les composantes d'une mammographie "Mammomat Inspiration" by Siemens.	8
1.3	Les différentes formes possibles d'une masse : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière [12]	9
1.4	Les différents contours possibles d'une masse : a) Circonscrit, b) Microlobulé, c) Masqué, d) Indistinct et e) Spiculé [12].	10
1.5	Densité mammaire selon le lexique BI-RADS : a) Densité < 25%, b) Densité entre 25 et 50%, c) Densité entre 50 et 75% et d) Densité > 75% [12].	10
1.6	Principe des SAD [Sprague, 1987] [4].	12
2.1	a) Image originale, b) Image segmentée, c) tumeur extrait après MGMRT et ROI [19].	14
2.2	a) Mammographie numérisée originale, b) Région du sein, c) Maxima locale (indiqués par points sombres), d) Résultat de la segmentation après une réduction de faux positif [20].	15
2.3	Résultats expérimentaux de la méthode proposée. a) Image originale de la base de données mini-MIAS, (b) image prétraitée, (c) Image en gradient, (d) Transformation du bassin versant, (e) Ouverture, (f) Clôture, (g) Reconstitue de l'ouverture-fermeture, (h) Masse identifiée.	16
2.4	Illustration des étapes de la CAO en mammographie pour le système fonctionnant à environ 1 FP/image. a) mammographie DDSM originale avec l'annotation fournie par le DDSM d'une masse présente, b) la sortie de la carte de probabilité du modèle correspondant à l'étape CAO, c) Marque CAO générée lors de l'étape finale [23].	17
3.1	Schéma de toutes les étapes de traitement utilisées dans notre algorithme.	19
3.2	Hibiscus et son histogramme [28].	20
3.3	Fonction idéale d'égalisation d'un histogramme [28].	21
3.4	Égalisation d'histogramme [28].	21
3.5	Fonctions de seuillage et « fenêtre d'intensité » [28].	22
3.6	Seuillage [28].	23
3.7	Textures. a) Fourrure de chien, b) Herbe, c) Galets de rivière, d) Liège, e) Textile à carreaux, f) Tissu tricoté [30].	24

3.8	Exemple d'une forêt aléatoire de 600 arbre [38].	28
4.1	Exemples des images dans INbreast [39].	31
4.2	Version demo enligne de Jupyter notebook [40].	32
4.3	Les logos des bibliothèques utilisées.	33
4.4	Diagramme de notre système de classification des images mam- mographie.	34
4.5	Résultat du pré-traitement.	35
4.6	L'image mammographie après la méthode de seuillage (binaire).	35
4.7	La suppression des régions non unitives après le zoom d'image : a)seuillage , b)morphologie, et c) reconstruction	36
4.8	Courbe de ROC pour le RF	37

Liste des tableaux

1.1	Critères de distinction entre une tumeur bénigne et une autre maligne [10].	7
2.1	Comparaison entre différentes méthodes de classification des masses mammaires	18
3.1	Les caractéristiques de texture de GLCM utilisées dans notre étude [toutes les équations sont de [33] sauf l'équation de dissimilarité (ou FDM) est de [34]].	26
4.1	La répartition des données utilisées.	31
4.2	Les résultats des différentes méthodes de classification.	36

Acronymes

ACC : Accuracy.
ACR : American College of Radiology.
ASM : Angular Second Moment.
BI-RADS : Breast Imaging Reporting and Data System.
BLHE : Band Limited Histogram Equalization.
CAO : Computer Aided design.
CLAHE : Contrast Limited Adaptive Histogram Equalization.
CPU : Central Processing Unit.
CVM : Core Vector Machine.
DAO : Diagnostique assisté par ordinateur.
DDR4 : Double Data Rate 4.
DDSM : Digital Database for Screening Mammography.
FDM : First-order element Difference Moment.
FPR : False Positive Reduction.
GB : Gigabyte.
GEN : Generation.
GHz : Gigahertz.
GLCM : Gray Level Co-occurrence Matrix.
GO : Gigaoctet.
GPU : Graphics Processing Unit.
HP : Hewlett-Packard (company).
IDE : Integrated Development Environment.
IDM : Inverse Difference Moment.
JPG : ou bien JPEG qui signifie : Joint Photographic Experts Group.
KNN : K-Nearest Neighbors.
LBP : Local Binary Pattern.
LPE : Ligne de Partage des Eaux.
LTEM : Laws' Texture Energy Measures.
MCC : Microcalcifications.
MGMRGT : Modified Gradient Magnitude Region Growing Technique.
MHz : Megahertz.
mini-MIAS : Se réfère à une version réduite de la base de données MIAS des images mammographiques tandis que MIAS signifie : Mammographic Image Analysis Society.
MO : Megaoctet.
NA : Not Available.
NumPy : Numerical Python.
OpenCV : Open Computer Vision.

RAM : Random Access Memory.

RF : Random Forests.

RN : Réseaux de neurones.

ROC : Receiver Operating Characteristic.

ROI : Regions of Interest .

RPM : Revolutions Per Minute.

SAD : Système d'aide à la décision.

SATA : Serial Advanced Technology Attachment.

SciPy : Scientific Python.

SSD : Solid-State Drive.

SVM : Support Vector Machine.

Introduction

Le cancer du sein est le cancer le plus fréquent chez la femme dans le monde. Plusieurs rapports sur l'épidémiologie et la gravité du cancer du sein sont publiés par différentes organisations [1].

Le cancer du sein a désormais dépassé le cancer du poumon en tant que cancer le plus souvent diagnostiqué dans le monde, selon les statistiques publiées par le Centre international de recherche sur le cancer (CIRC) en décembre 2020.

Au cours des deux dernières décennies, le nombre total de personnes diagnostiquées avec un cancer a presque doublé, passant d'environ 10 millions en 2000 à 19,3 millions en 2020. Aujourd'hui, une personne sur cinq dans le monde développera un cancer au cours de sa vie. Les projections suggèrent que le nombre de personnes recevant un diagnostic de cancer augmentera encore dans les années à venir et sera près de 50% plus élevé en 2040 qu'en 2020.

Le nombre de décès par cancer a également augmenté, passant de 6,2 millions en 2000 à 10 millions en 2020. Plus d'un décès sur six est dû au cancer.[2].

Parmi diverses modalités, la mammographie est la méthode la plus utilisée pour dépister différentes anomalies dans les seins. Toutefois, la mammographie n'est pas un examen très précis ni affirmatif, car il présente plusieurs imperfections. Les images obtenues par cette technique ne sont pas toujours claires, ce qui amène les radiologues dans plusieurs occasions à des erreurs de diagnostic qui se divisent en deux groupes : les faux-positifs et les faux négatifs.

Au cours des deux dernières décennies, de nombreux scientifiques ont tenté d'aider les radiologues à détecter et à diagnostiquer ces anomalies. Des techniques dédiées au traitement des images de la mammographie, ainsi que des systèmes de diagnostic assisté par ordinateur (DAO) sont développés pour cette fin.

La tendance actuelle réside sur la collection et l'analyse des données obtenues à partir des images numériques de la mammographie afin d'élaborer des systèmes intelligents qui contribuent à l'augmentation de la qualité du diagnostic. Il est cependant important de noter que les systèmes de diagnostic assisté par ordinateur (DAO) ou système d'aide au diagnostic sont conçus pour assister le radiologue seulement comme une seconde interprétation et jamais comme substitut.

Les masses et les micro-calcifications (MCC) sont les deux constatations les plus

fréquentes en mammographie. La détection des masses est plus difficile que la détection des MCC car les caractéristiques de masse peuvent être ambiguës ou similaire au parenchyme des seins. Les masses sont généralement situées dans les régions denses du sein. De plus, ils ont des limites plus lisses que les MCC et des formes plus variées.

La plupart des systèmes de DAO commerciaux pour détecter les MCC ont atteint 100% du taux de détection, mais la détection le taux de masse est toujours inférieure à 90% [3].

Ces facteurs font un défi à la fois pour les humains (radiologues) et machines (systèmes DAO). Il est rapporté que la plupart des anomalies manquées par les radiologues sont liées aux masses cancéreuses.

Le but de ce projet est d'utiliser des méthodes et des algorithmes efficaces pour augmenter la sensibilité des systèmes de détection dans le but de réduire les faux-négatifs.

L'objectif de ce travail est de mettre au point un système d'aide au diagnostic pour la détection et la classification des masses dans des images mammographies. Ce système est basé sur l'utilisation, en premier lieu, d'une méthode de segmentation qui se base sur la technique de seuillage, puis une phase de caractérisation est effectuée, à la fin, une classification est réalisée en utilisant le classifieur de type Random Forest.

Ce manuscrit est divisé en plusieurs chapitres. Tout d'abord, on a collecté toutes les informations qui servent à nous éclaircir la problématique, à savoir l'anatomie du sein, la définition de la mammographie et du DAO, les différentes méthodes de classification des images etc. Ces informations construisent le cadre théorique de notre projet qui est l'objectif du premier chapitre.

Dans le deuxième chapitre, on a cité les travaux de la littérature qui ont utilisé les différentes techniques de traitement de l'image pour traiter les images de la mammographie.

Le troisième chapitre constitue notre proposition, dans laquelle on explique le principe de notre application et les différentes parties de notre proposition.

Dans le quatrième chapitre, on discute les résultats obtenus en appliquant notre algorithme sur des clichés mammaires numériques. On termine ce travail par une conclusion générale.

1

Chapitre 01 : Informations générales

1 Introduction

La compréhension de la procédure d'interprétation de l'imagerie mammaire par les radiologues est cruciale dans le processus d'identification des verrous qui semblent représenter un frein à un dépistage efficace du cancer du sein par les experts.

Cela passe par une étude de l'aspect médical du cancer du sein, dont le but est de nous aider à nous approprier l'expertise développée en imagerie du sein par les radiologues. Ainsi, dans l'idée d'aider les radiologues à améliorer la précision de leur diagnostic tout en réduisant le nombre de biopsies inutiles, les chercheurs en vision par ordinateur ont proposé des systèmes de diagnostic assisté par ordinateur dont le but est d'identifier les lésions mammaires et de les classer de façon automatique.

Malgré la réelle difficulté posée par la qualité des images, la conception de ces systèmes ne cesse de connaître de grandes avancées depuis plus d'une vingtaine d'années. Notamment, en raison de leur capacité à fournir un résultat objectif et reproductible, donc fiable ce qui n'est toujours pas le cas des radiologues [4].

Dans ce chapitre, nous présenterons quelque définition sur l'anatomie de sein la définition de cancer du sein chez la femme, puis la définition de système d'aide au diagnostic et dérivant méthodes d'aide au diagnostic.

2 L'anatomie du sein

Le sein est composé d'une glande mammaire, de fibres de soutien (ligaments de Cooper) et de graisse (tissu adipeux), le tout est recouvert par la peau.

La quantité de chacune de ses composantes peut varier d'une femme à l'autre. Le sein est situé par-dessus le muscle pectoral. On trouve également dans le sein des

nerfs, des vaisseaux sanguins et lymphatiques.

La glande mammaire est divisée en 15 à 20 sections qu'on appelle lobes, composés de lobulés. Ceux-ci sont reliés à des canaux qui se rendent sous le mamelon (situé au centre du sein). On peut également observer des chaînes de ganglions lymphatiques qui filtrent les microbes et protègent le corps contre l'infection et la maladie.

Le cancer du sein peut se développer tant au niveau d'un canal galactophore que d'un lobule et il peut également se retrouver au niveau des ganglions lymphatiques (figure 1.1) [5].

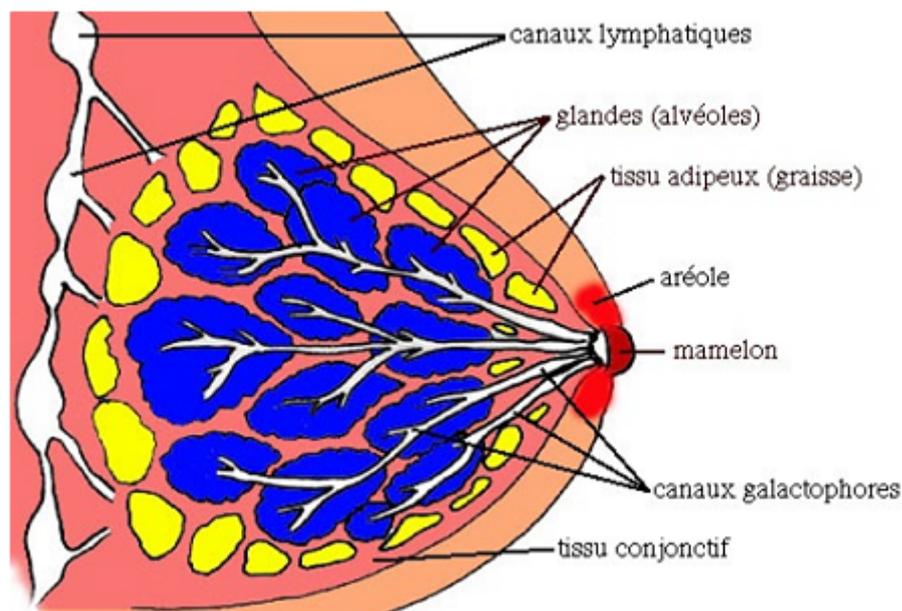


FIGURE 1.1 – Une illustration présente l'anatomie du sein [6].

3 Qu'est-ce que le cancer du sein ?

Le cancer du sein prend naissance dans les cellules du sein. La tumeur cancéreuse (maligne) est un groupe de cellules cancéreuses qui peuvent envahir et détruire le tissu voisin. Elle peut aussi se propager (métastases) à d'autres parties du corps.

Les cellules du sein subissent parfois des changements qui rendent leur mode de croissance ou leur comportement anormal. Ces changements peuvent engendrer des affections non cancéreuses (bénignes) du sein, comme l'hyperplasie atypique et des kystes. Ils peuvent aussi entraîner la formation de tumeurs non cancéreuses, dont les papillomes "intra-canalaire".

Dans certains cas, cependant, des modifications dans les cellules mammaires peuvent causer un cancer du sein.

Le cancer du sein apparaît le plus souvent dans les cellules tapissant les canaux, qui sont des tubes qui transportent le lait des glandes au mamelon. Ce type de cancer du sein est appelé carcinome canalaire.

Le cancer peut aussi se former dans les cellules des lobules, qui sont les groupes de glandes productrices de lait. Ce type de cancer porte le nom de carcinome lobulaire.

Le carcinome canalaire et le carcinome lobulaire peuvent être in situ, c'est-à-dire que le cancer reste dans son emplacement d'origine et qu'il n'a pas envahi les tissus voisins. Ils peuvent également être infiltrant, ou invasifs, c'est-à-dire qu'ils ont envahi les tissus voisins.

Des types de cancer du sein moins fréquents peuvent aussi se manifester. Le cancer inflammatoire du sein, la maladie de Paget du sein et le cancer du sein triple négatif en sont des exemples.

Des types rares de cancer du sein sont entre autres le lymphome non hodgkinien et le sarcome des tissus mous [7].

4 Les différents types de tumeurs

Il existe plusieurs types de tumeurs avec une variété de noms. Leurs noms reflètent généralement leurs formes, l'origine de la cellule et le type de tissu dans laquelle elles apparaissent.

En général, les tumeurs sont divisées en trois groupes [8] :

- Bénignes
- Prémalignes
- Malignes

Le tableau 1.1 décrit les points par lesquels on peut différencier une tumeur bénigne d'une autre maligne.

4.1 Tumeurs bénignes

Ces tumeurs ne sont pas cancéreuses et ne peuvent pas se propager. Une tumeur bénigne restera dans sa forme actuelle. La tumeur ne revient généralement pas après avoir été enlevée.

Même si elles ne sont pas nocives pour la santé humaine, certaines peuvent exercer des pressions sur les nerfs et les vaisseaux sanguins ce qui cause des douleurs gênantes et d'autres effets négatifs [8].

On désigne sous le nom de mastopathies les principales maladies bénignes du sein. On retrouve [9] :

- **Kystes**
Les kystes sont des cavités remplies de liquide. Uniques ou multiples, ils peuvent apparaître et augmenter de taille rapidement.
Le diagnostic se fait par l'intermédiaire d'une échographie. Une ponction peut être pratiquée, plutôt en présence de douleurs.
- **Adénofibromes ou fibroadénomes**
Ce sont des nodules de forme arrondie, fermes, uniques, bien délimités, indolores ; ils roulent sous les doigts.

Ils augmentent souvent de taille en période prémenstruelle. Ils peuvent diminuer voire disparaître tout seul. Ils se transforment rarement en cancer. Ils touchent souvent la femme de moins de 30 ans.

- **Mastite inflammatoire**

La mastite une maladie inflammatoire du sein que l'on peut rencontrer au cours de l'allaitement. On l'appelle aussi mastite lactationnelle.

Dans les formes les plus sévères, elle peut évoluer vers l'abcès

- **Abcès**

Il s'agit d'une collection de pus qui se présente sous la forme d'une bosse sensible et rouge.

L'abcès peut également se former en dehors de la grossesse ou de l'allaitement. Dans ce cas, il se situe dans les canaux sous l'aréole entourant le mamelon et a tendance à récidiver.

En plus des antibiotiques, le traitement consiste à inciser et mécher la plaie. Une ablation des canaux peut être envisagée lorsque les récurrences sont trop nombreuses.

- **Papillome intra canalaire**

Il s'agit d'une tumeur bénigne qui se développe à l'intérieur d'un canal galactophore, responsable d'un écoulement jaunâtre parfois sanguinolent. On peut la voir à l'échographie. Son exérèse est recommandée.

4.2 Tumeurs pré-malignes

Une tumeur pré-maligne n'est pas encore Cancéreuse mais semble développer les propriétés du cancer. Ce type des tumeurs nécessite une surveillance étroite [8].

4.3 Tumeurs malignes (Cancéreuses)

On les appelle aussi cancer du sein. Le cancer peut se développer à partir d'un canal galactophore ou d'un lobule. Lorsqu'il se situe au niveau des cellules de la glande mammaire on parle d'adénocarcinome.

C'est le cancer le plus fréquent chez la femme et la principale cause de mortalité entre 35 et 65 ans.

Actuellement une femme sur huit est touchée et dans les 3/4 des cas, il se développe après 50 ans. Il est influencé par les hormones féminines [9].

4.4 Tumeurs phylodes

Également appelée cysto-sarcome phyllode, cette tumeur est rare (2%). Elle se présente sous une forme bosselée qui peut rapidement augmenter de volume entraînant alors une déformation du sein.

Cette tumeur est généralement bénigne. Dans certains cas elle peut se comporter comme une tumeur cancéreuse et dans moins de 5% des cas elle peut être cancéreuse [9].

Tumeurs bénigne	Tumeur maligne
Bien limitée	Mal limitée
Encapsulée	Non encapsulée
Histologiquement semblable au tissu d'origine	Plus ou moins semblable au tissu d'origine (dédifférenciation, différenciation aberrante)
Cellules régulières	Cellules irrégulières (cellules cancéreuses)
Croissance lente	Croissance rapide
Refoulement sans destruction des tissus voisins	Envahissement des tissus voisins
Pas de récurrence locale après exérèse complète	Récurrence possible après exérèse supposée totale
Pas de métastase	Métastase(s)

TABLE 1.1 – Critères de distinction entre une tumeur bénigne et une autre maligne [10].

5 Définition de la mammographie

La mammographie, aussi appelée radiographie des seins, est l'examen le plus efficace pour dépister le plus tôt possible un cancer du sein, même lorsqu'il n'y a pas de symptômes.

Elle permet de visualiser les structures à l'intérieur du sein et de détecter les lésions de petite taille susceptibles de ne pas être repérées par palpation (kystes, calcifications et tumeurs).

C'est un examen d'une durée de quelques minutes qui est réalisé par une technologue en radiologie. Il consiste à prendre un minimum de deux radiographies de chaque sein, une horizontale et une en oblique (de côté) en appliquant une compression sur le sein pendant 10 à 15 secondes afin d'obtenir une image de qualité et réduire la quantité de radiations émises.

Il n'y a aucun risque de traumatisme ou de blessure, mais la compression peut causer un certain inconfort, parfois même une légère douleur chez certaines personnes [11].

La figure 1.2 ci-dessous illustre les composantes principales d'une mammographie de type "Mammomat Inspiration (Siemens)".

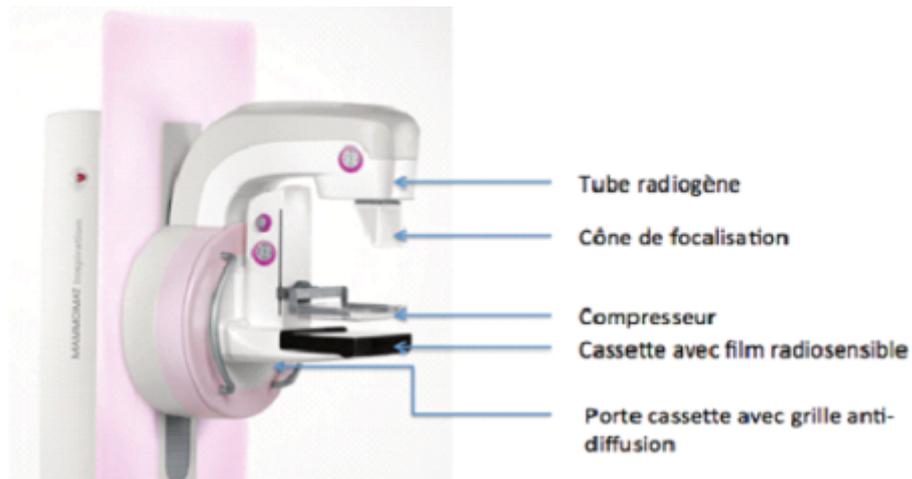


FIGURE 1.2 – Les composantes d’une mammographie "Mammomat Inspiration" by Siemens.

6 Les masses

Une opacité ou une masse est une zone occupant un espace important et vue sous deux angles différents. Si une opacité potentielle est observée sur une seule incidence, elle est appelée asymétrique jusqu’à ce que son caractère tridimensionnel soit confirmé.

Différentes caractéristiques de ces masses sont à écrire à savoir la forme, le contour et la densité [12].

6.1 La forme

Selon la description du BI-RADS (D’Orsi et al., 2003 [13]), les masses mammaires peuvent avoir la forme ronde (figure 1.3 a), ovale (figure 1.3 b), lobulée (figure 1.3 c) ou irrégulière (figure 1.3 d) [12].

- (a) **Ronde** : Il s’agit de masse sphérique, circulaire ou globuleuse.
- (b) **Ovale** : Elle présente une forme elliptique (ou en forme d’œuf).
- (c) **Lobulée** : La forme de la masse présente une légère ondulation.
- (d) **Irrégulière** : Cette appellation est réservée aux masses dont la forme est aléatoire et ne peut être caractérisée par les termes cités ci-dessus.

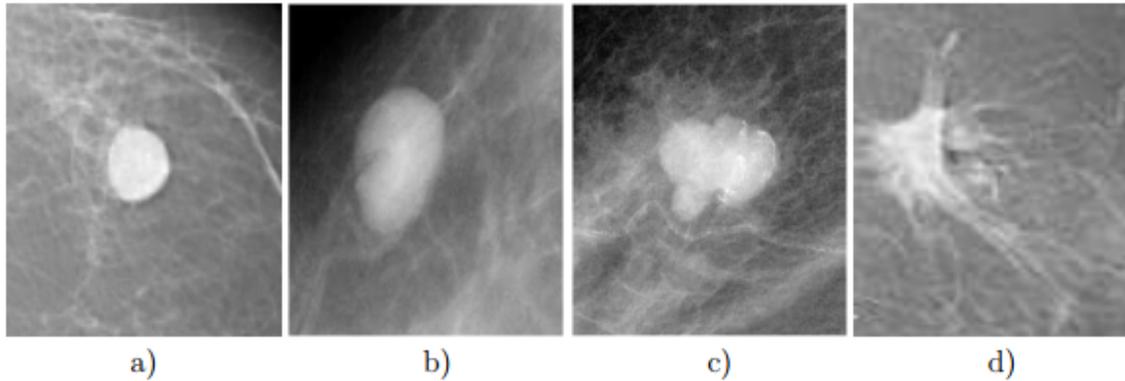


FIGURE 1.3 – Les différentes formes possibles d’une masse : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière [12]

6.2 Le contour

Le contour des masses mammaires est soit circonscrit (figure 1.4 a), soit microlobulé (figure 1.4 b), soit masqué (figure 1.4 c) soit indistinct (figure 1.4 d), soit spiculé (figure 1.4 e). On détaille dans ce qui suit ces différentes notions : [12]

- (a) **Circonscrit** : Il s’agit d’une transition brusque entre la lésion et le tissu environnant. Le contour est alors net et bien défini. Pour qu’une masse soit qualifiée de circonscrite, il faut qu’au moins 75% de son contour soit nettement délimité.
- (b) **Microlobulé** : Dans ce cas, de courtes dentelures du contour créent de petites ondulations.
- (c) **Masqué** : Un contour masqué est un contour qui est caché par le tissu normal adjacent. Ce terme est employé pour caractériser une masse circonscrite dont une partie du contour est cachée.
- (d) **Indistinct** : Dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre ‘a une infiltration.
- (e) **Spiculé** : La masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.

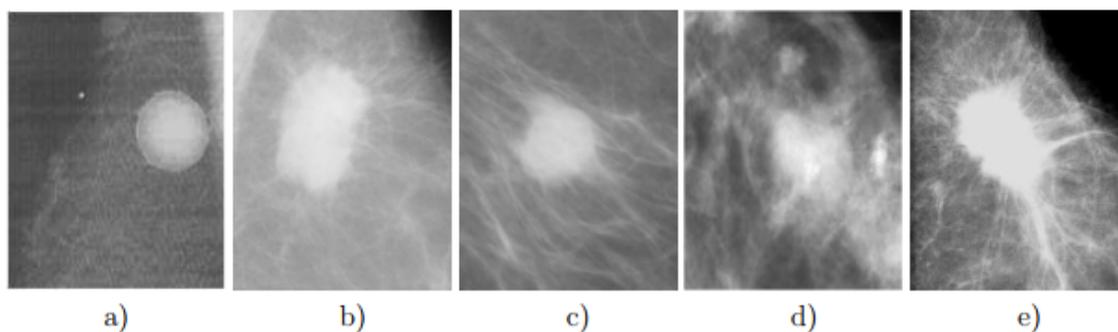


FIGURE 1.4 – Les différents contours possibles d’une masse : a) Circonscrit, b) Microlobulé, c) Masqué, d) Indistinct et e) Spiculé [12].

6.3 La densité

L’apparence du sein normal varie fortement d’une femme à l’autre. Le facteur le plus remarquable est la grande variabilité de la densité radiologique de la zone mammaire.

Wolfe a été le premier à établir une relation entre la densité du tissu mammaire et le risque de développer un cancer (Wolfe, 1976 [14]). D’autres études remettent en question ce lien (LeTretout et Dilhuydy, 1988 [15]). Néanmoins, cette classification tissulaire est utilisée dans d’autres travaux sans faire de lien avec le facteur de risque (Suckling et al., 1995 [16]).

Afin de standardiser les rapports mammographiques, la classification BI-RADS de l’ACR définit 4 classes de la composition du sein [12].

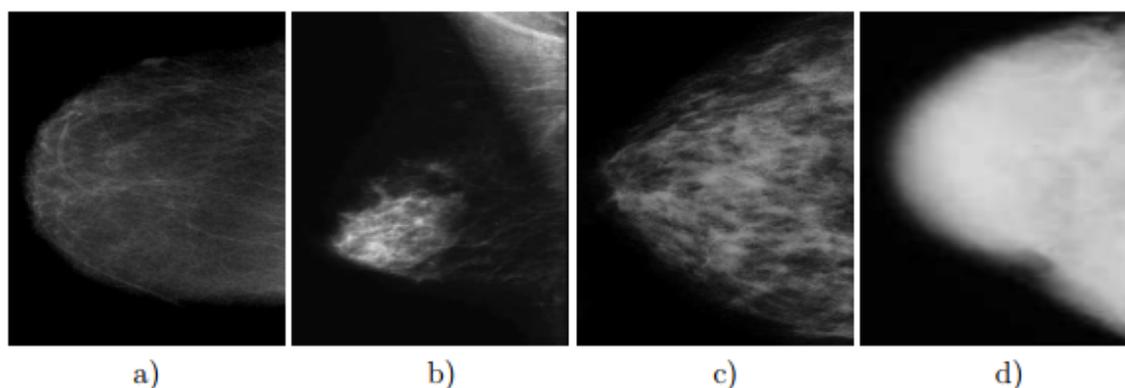


FIGURE 1.5 – Densité mammaire selon le lexique BI-RADS : a) Densité < 25%, b) Densité entre 25 et 50%, c) Densité entre 50 et 75% et d) Densité > 75% [12].

- (a) **Stade 1** : Le sein est presque entièrement graisseux et homogène, radio transparent et facile à lire (moins de 25% de la glande mammaire).
- (b) **Stade 2** : Il y a des opacités fibro-glandulaires dispersées. Le sein est graisseux et hétérogène (approximativement 25 à 50% de la glande mammaire).
- (c) **Stade 3** : Le tissu mammaire est dense et hétérogène (approximativement 51 à 75% de la glande mammaire).

- (d) **Stade 4** : Le tissu mammaire est extrêmement dense et homogène. La mammographie est alors difficile à interpréter puisque la densité peut masquer une lésion (plus de 75% de la glande mammaire) (figure 1.5).

7 Qu'est-ce qu'un SAD ?

Un SAD est un outil permettant d'assister le décideur dans sa prise de décision en effet, un processus de décision ne peut pas être entièrement automatisable [Pomerol, 1992].

Un SAD est un système de traitement de l'information qui permet d'extraire et de donner au décideur l'information nécessaire au processus de prise de décision [4].

7.1 Rôle d'un SAD

Un SAD est système informatique dont le rôle est d'assister le décideur tout au long du processus de décision.

Un processus de décision dans le cadre défini par la gestion des organisations, se compose de quatre phases [Lévine and Pomerol, 1989] [4] :

- Une phase d'information.
- Une phase de conception.
- Une phase de choix.
- Une phase d'évaluation du choix.

7.2 Principe d'un SAD

Un SAD se compose de trois modules [4] :

- Un module de dialogue.
- Un module contenant les données.
- Un module contenant les procédures de calcul ou modèles.

Comme l'indique la figure 1.6, le module de dialogue est interconnecté avec les deux autres modules. Il constitue l'interface entre l'utilisateur et le reste du système.

- **Le module de dialogue**

Un composant essentiel d'un SAD. Par l'intermédiaire des interfaces gérées par ce module, le décideur accède aux données et aux fonctions de calcul, et le système utilise le même vecteur pour lui communiquer le résultat des manipulations effectuées par le décideur.

- **Le module base de données**

Assure la fonction du mémoire, il stocke non seulement les données, de façon permanente ou temporaire, mais il gère également l'enregistrement de données volatiles ainsi que l'effacement de ces données selon le souhaite de l'utilisateur.

Les données volatiles correspondent aux résultats obtenus au cours de traitements des données.

- **Le module "modèle"**

Contient l'ensemble des procédures de calcul utilisées dans les différents traitements standards et procédure de représentation des données.

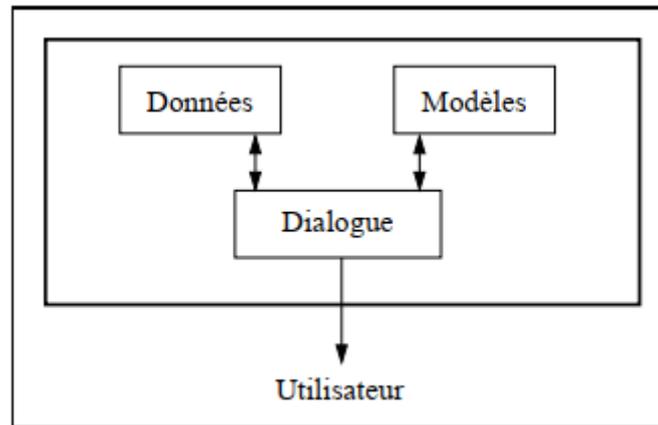


FIGURE 1.6 – Principe des SAD [Sprague, 1987] [4].

7.3 Différentes méthodes de l'aide au diagnostic

L'approche purement numérique de l'aide à la décision. Les analyses discriminantes et les méthodes statistiques multidimensionnelles.

Pour l'aide à la décision (diagnostic par exemple), ces méthodes sont appliquées sur un échantillon de cas (dont on connaît le diagnostic) pour déterminer la fonction discriminante.

Schématiquement, si l'on suppose que l'on cherche à discriminer les sujets malades des non malades.

On peut dire qu'il s'agit, dans un espace à dimensions correspondant aux variables décrivant les patients, de trouver le plan qui sépare au mieux (au sens d'un certain critère) les points correspondant aux malades et les points correspondant aux sujets sains, cette fonction obtenue est testée sur un autre échantillon de données afin d'évaluer sa validité.

8 Conclusion

A travers ce chapitre, nous avons cité plusieurs notions théoriques qui construisent le cadre théorique de notre projet.

Dans le chapitre qui suit, nous abordons les différentes techniques de traitement et de classification des images mammaire. Ainsi que les travaux effectués dans ce domaine.

2

Chapitre 02 : L'état de L'art

Grâce aux progrès de la technologie informatique, les radiologues sont en mesure d'améliorer leur interprétation des images en utilisant des capacités de calcul qui peuvent améliorer la qualité des images de mammographie.

Dans ce chapitre, nous allons étudier les différentes méthodes de CAO et tenter de les comparer. À partir d'une critique qui il fait par S. Sukumaran et al. [17], l'algorithme CAO dans le domaine de la classification des masses mammaire est basé sur cinq (05) méthodes principales [17] :

1. Pré-traitement

C'est la première méthode, les techniques de pré-traitement sont très nécessaires pour trouver l'orientation de la mammographie afin d'éliminer le bruit et d'améliorer la qualité de l'image.

2. Extraction de la région d'intérêt ROI

Dans cette étape il s'agit en fait d'une segmentation de l'image pour extraire la région des masses si elle se trouve avec une partition de l'en groupes de Pixels homogènes avec l'image par rapport à certains critères.

Ces critères peuvent inclure des valeurs d'intensité, de texture, de couleur, de gamme, de normale de surface et de courbure de surface.

3. Extraction de caractéristiques

L'extraction de caractéristiques peut être définie comme une mesure ou une analyse quantitative d'images médicales.

De nombreux types de caractéristiques peuvent être extraits pour traiter les anomalies des mammographies.

4. Techniques de regroupement

Le regroupement est un processus de partitionnement ou de regroupement d'un ensemble de motifs non étiquetés en un certain nombre de grappes dans lesquelles des motifs similaires sont attribués à une grappe. Chaque motif peut être représenté par un vecteur avec de nombreux paramètres ou attributs.

5. Classification

Une fois que les caractéristiques liées à la masse sont extraites et sélectionnées, les caractéristiques sont entrées dans un classifieur pour classer les zones suspectées.

1 Segmentation des masses mammaires

SIM. Ali et al. [18], ils ont présenté un algorithme qui aide les radiologues à identifier les tumeurs mammaires à leurs débuts.

Cet algorithme élimine automatiquement les artefacts 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. La détection de la région d'intérêt est réalisée en utilisant la méthode de seuillage d'Otsu.

R. Dubey et al. [19], ils font une comparaison entre la technique de Modified gradient magnitude region growing technique (MGMRGT) (figure 2.1) et l'algorithme de segmentation des bassins versants "watershed segmentation", l'algorithme de segmentation des bassins versants c'est une méthode manuelle par contre la méthode de croissance de la région à gradient de magnitude modifié MGMRGT, elle permet de détecter la tumeur nés pas d'une façon totalement automatiquement mais semi-automatique.



FIGURE 2.1 – a) Image originale, b) Image segmentée, c) tumeur extrait après MGMRGT et ROI [19].

Comme les mammographies contenant des lésions sont généralement caractérisées par un faible contraste, beaucoup d'objets tissulaires normaux sont inclus comme lésions suspectées. Afin de se débarrasser de ces régions tissulaires normales pour dégager ces tissus normaux, H. Zhang et al. [20] proposent une méthode appelle réduction de faux positive FPR, cette méthode est basée sur "Region Growing" avec un seuil adaptatif (figure 2.2).

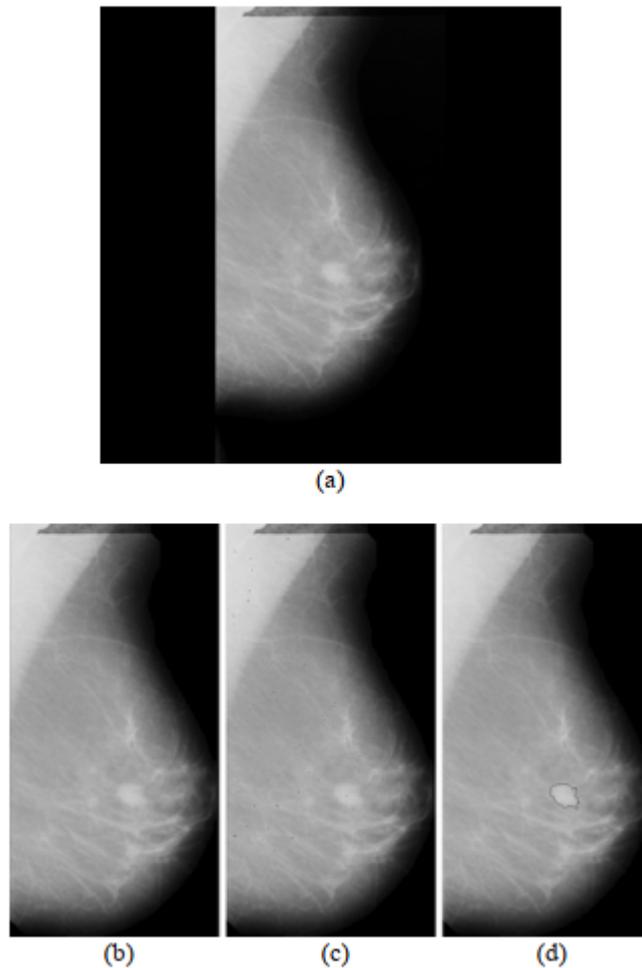


FIGURE 2.2 – a) Mammographie numérisée originale, b) Région du sein, c) Maxima locale (indiqués par points sombres), d) Résultat de la segmentation après une réduction de faux positif [20].

Samir M. Badawy et al. [21] proposent une méthode de segmentation à double seuil qui peut être illustrée simplement en choisissant deux valeurs de pixel L et U de notre image, où L est la limite de seuil inférieure et U est la limite de seuil supérieur, et en appliquant l'opération de seuillage comme suit :

Un pixel devient blanc si son niveau de gris est compris entre L et U , et noir si son niveau de gris est différent. Le résultat est un binaire (noir et blanc), le blanc pour tous les niveaux de gris pixel est entre les deux limites, les L , U et valeurs noires pour tous les autres.

Aziz Makandar et al. [22] ont proposé un travail de segmentation basée sur un seuil et il est modifié pour extraire le retour sur investissement avec la transformation du bassin versant "watershed transform" et l'opération morphologique. Ce travail peut être une bonne approche pour l'amélioration de la qualité de l'image dans le diagnostic et permet également de déterminer les caractéristiques relatives permettant de classer la tumeur comme bénigne ou maligne (figure 2.3).

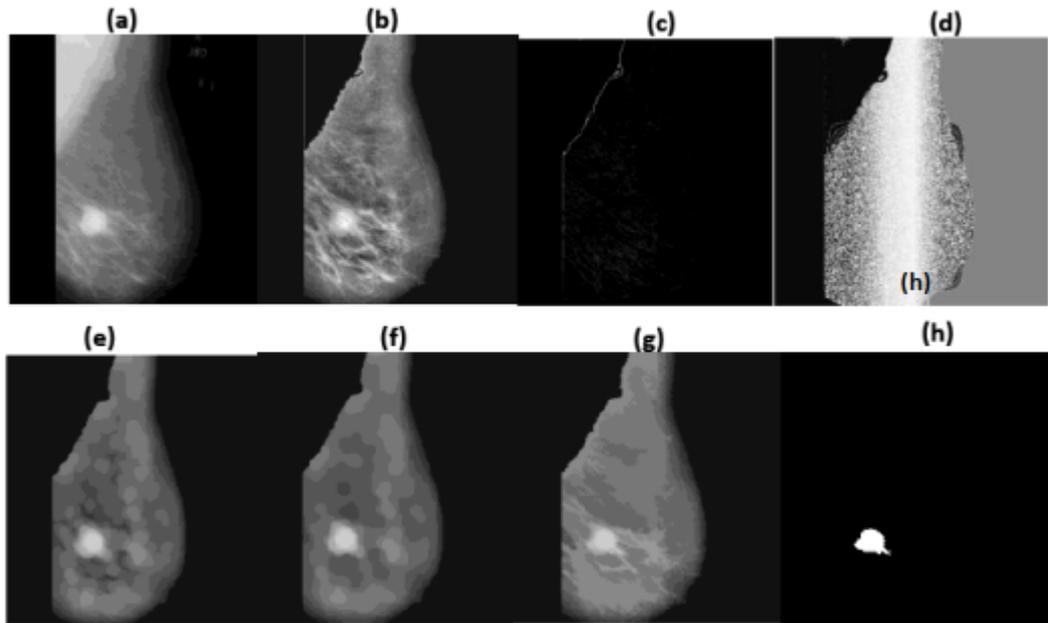


FIGURE 2.3 – Résultats expérimentaux de la méthode proposée. a) Image originale de la base de données mini-MIAS, (b) image prétraitée, (c) Image en gradient, (d) Transformation du bassin versant, (e) Ouverture, (f) Clôture, (g) Reconstitue de l'ouverture-fermeture, (h) Masse identifiée.

MA. Mazurowski et al. [23], ils ont présenté un système de CAO qui ils utilisé la segmentation du Template après le prétraitement des images mammographies (figure 2.4).

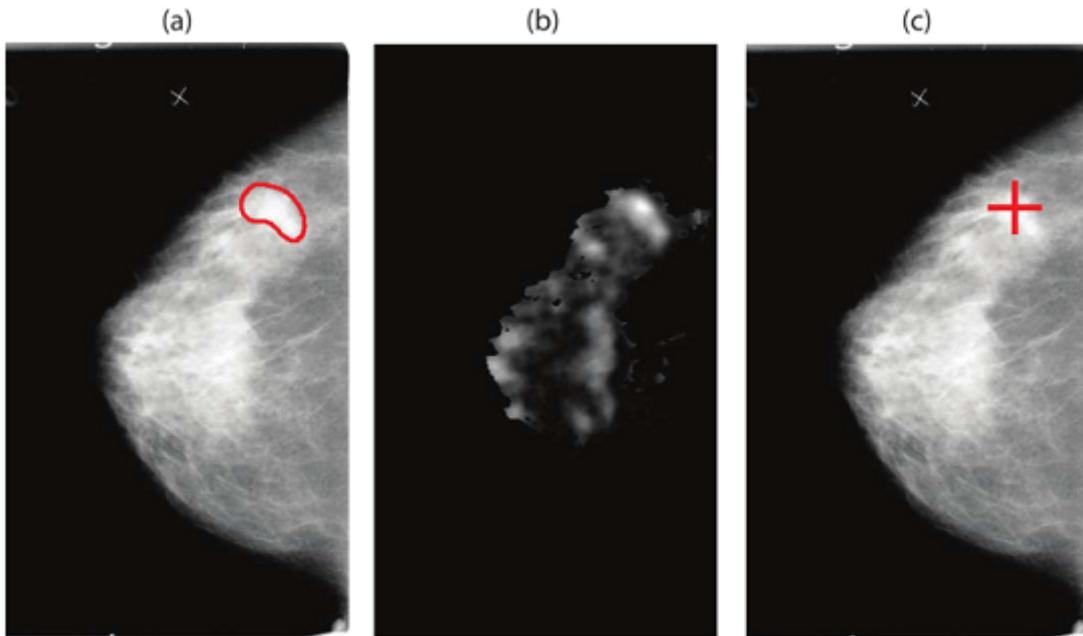


FIGURE 2.4 – Illustration des étapes de la CAO en mammographie pour le système fonctionnant à environ 1 FP/image. a) mammographie DDSM originale avec l'annotation fournie par le DDSM d'une masse présente, b) la sortie de la carte de probabilité du modèle correspondant à l'étape CAO, c) Marque CAO générée lors de l'étape finale [23].

Les auteurs pensent que le système de CAO a été très efficace dans la détection des masses malignes, avec leur flexibilité qui peut transférer avec succès les connaissances entre les différentes modalités offre de nombreux avantages en cette époque de progrès rapides dans le domaine du dépistage du cancer du sein.

2 Classification des masses mammaires

MM. Eltoukhy et al. [24], ils ont proposé une méthode d'Otsu pour la segmentation avec une réduction des faux positives. Les caractéristiques sont extraites par les deux méthodes : Curvelete et Wavelet.

Pour la classification, les auteurs ont utilisé le classificateur du voisin le plus proche KNN et la machine à vecteur de support SVM. Le but de cette méthode est de différencier les masses du tissu normal et ensuite de faire la distinction entre les masses bénignes et malignes.

JJ. Sonia et al [25] ont utilisé les mêmes méthodes de prétraitement et de segmentation que nous, mais pas la même méthode d'extraction des caractéristiques. Ils ont utilisé la mesure de l'énergie de texture (LTEM) et pour la classification le SVM. Ils ont conclu à partir de l'analyse des performances que le classifieur SVM est plus performant que les autres.

HA. Khan et al. [26] ont utilisé la méthode LBP invariante par rotation combi-

née avec le modèle uniforme pour l'extraction des caractéristiques et la méthode SVM pour la classification ; c'est une nouvelle méthodologie dans la classification de masse.

RF. dos Santos Teixeira et al. [27] ont utilisé la réduction aux niveaux de gris. Les caractéristiques ont été calculées par la Matrice de cooccurrence et le SVM comme un classifieur. Ils ont conclu que lorsque l'image subit une petite quantité de prétraitement, les résultats sont légèrement meilleurs que lorsque l'image originale est soumise directement au classifieur.

Auteurs	Prétraitements et segmentation	Extraction des caractéristiques	Classif.	ACC (%)
MM. Eltoukhy et al. [24]	False positive reduction + Otsu	Curvelet et wavelet	SVM/ KNN	96.66/ 93.37
JJ. Sonia et et al. [25]	Clahe + Otsu	Laws texture energy measures (LTEM)	CVM	95.89
HA. Khan et al. [26]	NA	LBP	SVM	74
RF. dos Santos Teixeira et al. [27]	Reduction grey level	GLCM	SVM	85

TABLE 2.1 – Comparaison entre différentes méthodes de classification des masses mammaires

3 Conclusion

Après cette comparaison entre les différentes méthodes de segmentation et de classification, nous pouvons observer qu'il existe de multiples méthodologies de CAO et de classification qui visent le même but, à savoir la classification des masses.

Les principales étapes d'une classification des masses impliquent l'amélioration, l'extraction de caractéristiques, le regroupement et la classification. Chaque étape peut être mise en œuvre en utilisant un certain nombre de méthodes.

À partir du tableau 2.1, nous pouvons observer que les étapes de pré-traitement et de segmentation sont très importantes pour obtenir une bonne classification.

3

Chapitre 03 : Partie théorique

1 Introduction

Dans ce chapitre, on explique théoriquement les différentes étapes de *traitement des images mammographies* qu'on a suivi dans notre approche. Nous commençons par une *égalisation d'histogramme*, ensuite la *segmentation* en utilisant le *seuillage*, la troisième étape est réservée pour *l'extraction des caractéristiques*, et à la fin un apprentissage et une *classification* (figure. 3.1).

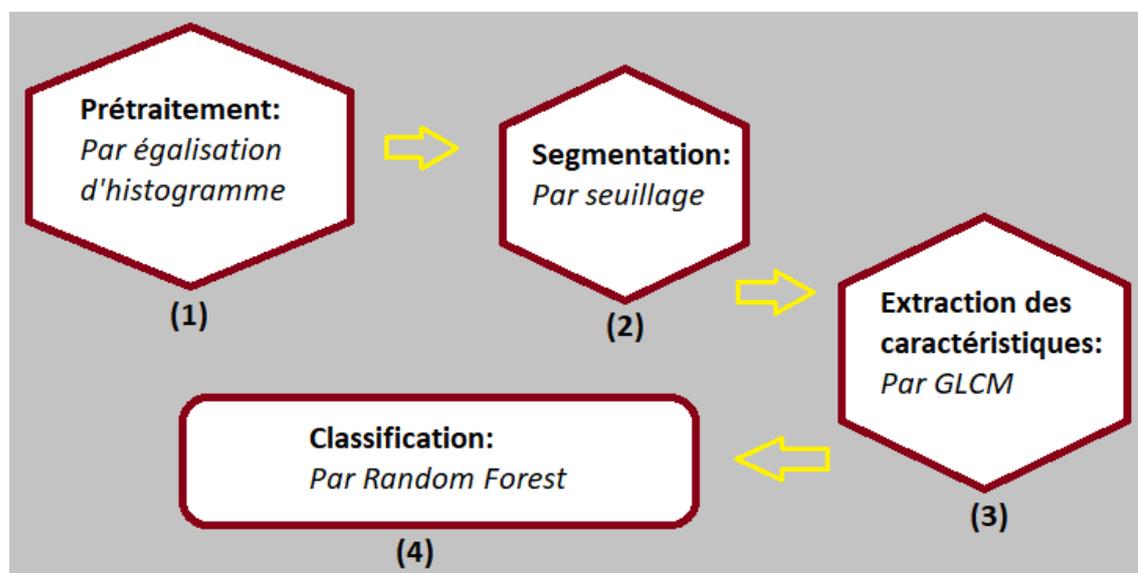


FIGURE 3.1 – Schéma de toutes les étapes de traitement utilisées dans notre algorithme.

2 Prétraitement : Égalisation de l'histogramme

2.1 Histogramme d'une image

Définition : L'histogramme d'une image est l'histogramme de la série de données correspondant aux niveaux de gris des pixels (figure. 3.2). C'est une fonction discrète définie par [28] :

$$\forall p \in \{0, \dots, 255\} \quad h_p = \text{Nombre des pixels ayant } p \text{ pour niveau de gris.}$$

On peut définir une version « continue » h de l'histogramme en faisant une interpolation (par exemple linéaire par morceaux) des valeurs h_p de sorte que :

$$\forall p \in \{0, \dots, 255\} \quad \mathbf{h}(p) = h_p$$

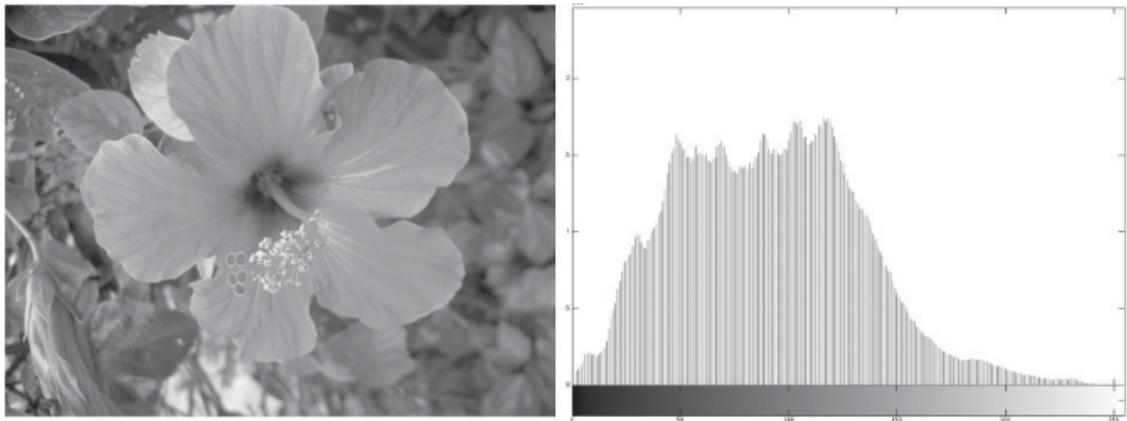


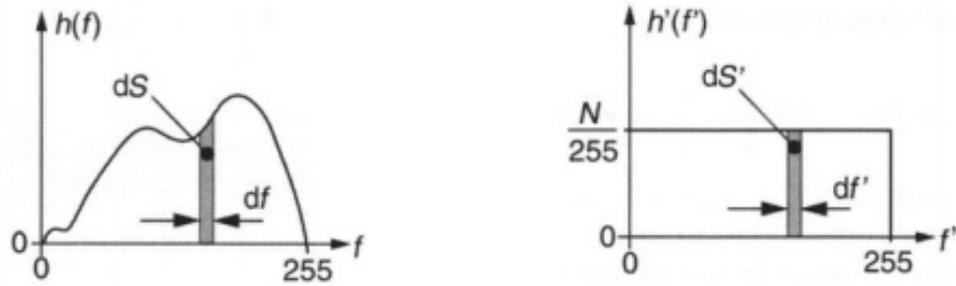
FIGURE 3.2 – Hibiscus et son histogramme [28].

2.2 Égalisation de l'histogramme

L'histogramme d'une image est rarement plat ce qui traduit une entropie non maximale. La transformation d'égalisation est construite de telle façon que l'histogramme de l'image transformée soit le plus plat possible (figure 3.3).

Cette technique améliore le contraste et permet d'augmenter artificiellement la clarté d'une image grâce à une meilleure répartition des intensités relatives (figure 3.4).

Considérons l'histogramme continu $f \mapsto \mathbf{h}(f)$. En notant $f' = t(f)$, l'histogramme égalisé $f' \mapsto \mathbf{h}'(f')$ doit s'approcher de la forme idéale décrite ci-dessous [28].



(a) Histogramme d'origine.

(b) Histogramme plat idéal.

FIGURE 3.3 – Fonction idéale d'égalisation d'un histogramme [28].

Deux surfaces élémentaires en correspondance dans les histogrammes initiaux et égalisés, présentent le même nombre de points ce qui permet d'écrire :

$$f' = t(f) = \frac{255}{N} \int_0^f h(s) ds$$

Ici, N est le nombre total de pixels de l'image et $\frac{h}{N}$ est l'histogramme normalisé (entre 0 et 1). En remplaçant l'intégration continue par une somme, on obtient la transformation d'égalisation discrète suivante :

$$f' = t(f) = \frac{255}{N} \sum_{i=0}^f h_i$$

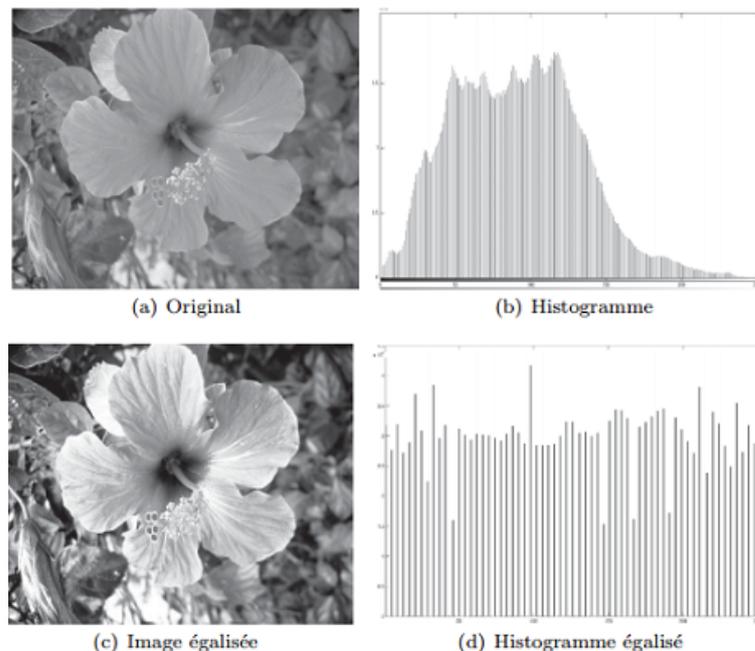


FIGURE 3.4 – Égalisation d'histogramme [28].

3 Segmentation : Seuillage binaire

3.1 Segmentation

La segmentation permet d'isoler certaines parties de l'image qui présentent une forte corrélation avec les objets contenus dans cette image, généralement dans l'optique d'un post-traitement. Les domaines d'application sont nombreux : médecine, géophysique, géologie, etc.

Dans le domaine médical, la segmentation d'images est extrêmement compliquée. En effet, pour chaque organe (cerveau, cœur, etc ...), l'approche est différente : l'outil de segmentation doit donc pouvoir s'adapter à un organe particulier, suivant une modalité d'acquisition particulière (scanners, radiographie, Imagerie par Résonance Magnétique, ...) et pour une séquence de données particulière.

En imagerie mathématique, on considère principalement deux types de segmentation [28] :

- La segmentation par **contours** qui permet de délimiter les différentes régions par leurs frontières.
- La segmentation par **régions** qui permet de caractériser les régions d'une image présentant une structure homogène et qui fait appel la plupart du temps à des outils statistiques.

3.2 Binarisation-seuillage

Le but de la binarisation d'une image est d'affecter un niveau uniforme aux pixels pertinents et d'éliminer les autres.

Le *seuillage* consiste à affecter le niveau 255 aux pixels dont la valeur est supérieure à un seuil S et le niveau 0 aux autres.

On peut avoir une approche plus « sélective » du seuillage en choisissant d'extraire une *fenêtre d'intensité* $[a, b]$. Avec cette transformation, la nouvelle image ne visualise que les pixels dont le niveau d'intensité appartient à l'intervalle $[a, b]$. Sous réserve d'une connaissance a priori de la distribution des niveaux de gris des objets de l'image originale, cette technique permet une segmentation d'objets particuliers de l'image. Le graphe de ces transformations est le suivant (figure 3.5) :

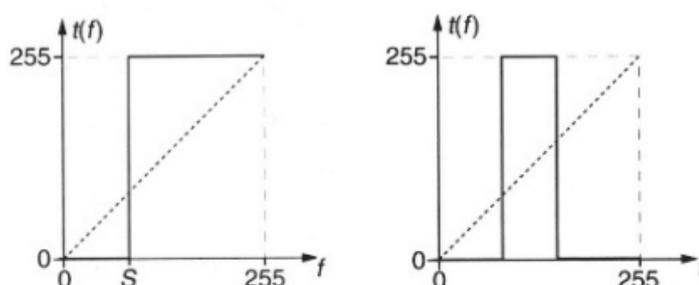


FIGURE 3.5 – Fonctions de seuillage et « fenêtre d'intensité » [28].

Ci-dessous un exemple montrant le résultat du seuillage avec et sans fenêtre (figure 3.6) :

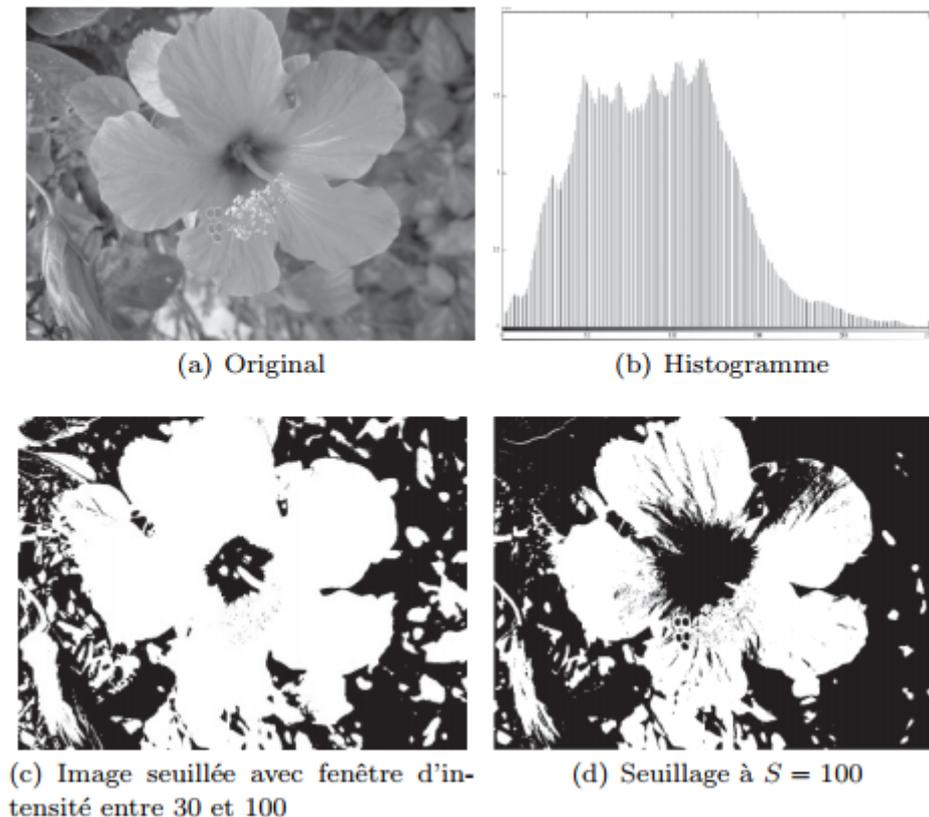


FIGURE 3.6 – Seuillage [28].

4 Extraction de caractéristiques : GLCM

4.1 Caractéristiques d'une image

Les caractéristiques d'image sont des descripteurs simplifiés de certains aspects d'une image, d'une partie de celle-ci ou objets dedans.

Un grand nombre de caractéristiques simples sont définies dans la littérature, mais ils appartiennent généralement à des catégories de base (forme, taille, luminosité et texture) [29].

4.2 Extraction des caractéristiques

L'extraction des caractéristiques d'image consiste à résumer/ convertir des données d'image brutes en représentations expressives qui sont plus informatives ou montrent une meilleure association avec un phénomène sous-jacent.

L'objectif d'une telle conversion est de mettre en évidence ou rendre explicites dans les données leurs éléments les plus pertinents par rapport à une tâche donnée [29].

4.3 La matrice de co-occurrence de niveau de gris (GLCM)

4.3.1 Texture

La texture contient des informations importantes sur la structure de la surface de l'image, qui est le facteur clé de la perception visuelle (figure 3.7).

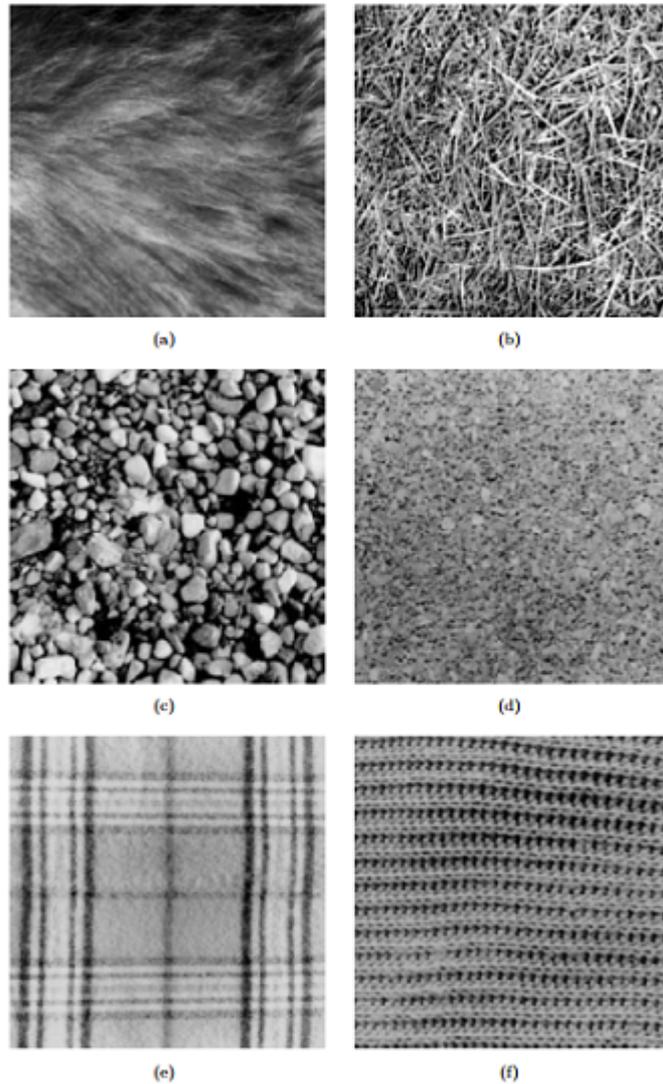


FIGURE 3.7 – Textures. a) Fourrure de chien, b) Herbe, c) Galets de rivière, d) Liège, e) Textile à carreaux, f) Tissu tricoté [30].

L'analyse de la texture est largement utilisée dans la classification des images, la reconstruction d'image en trois dimensions et la segmentation d'image. De nombreuses approches d'analyse de texture ont été développées, la *matrice de co-occurrence de niveau de gris (GLCM)* s'est avérée être l'une des approches d'analyse de texture les plus efficaces parmi les méthodes statistiques [31].

4.3.2 GLCM

Définition : La *matrice de co-occurrence* mesure la probabilité d'apparition des paires de valeurs de pixels situés à une certaine distance dans l'image. Cette probabilité représente le nombre de fois où un pixel de niveau de gris i apparaît à une distance relative d d'un pixel de niveau de gris j et selon une orientation $theta$ donnée. [Ben, 2014]

Haralick [32] a défini 14 indices à partir de cette matrice pour résumer les diverses propriétés de la texture, dans le but d'intégrer cette information dans une classification [31].

Dans notre cas, on a calculé 7 caractéristiques de texture de la matrice GLCM selon lesquels on construit notre base de données afin de procéder au classification des images (tableau 3.1) :

1. Contrast
2. Homogeneity or Inverse difference moment (IDM)
3. Energy or Angular Second Moment (ASM)
4. Entropy
5. Dissimilarity or First-order element difference moment (FDM)
6. Cluster shade
7. Cluster prominence

Principe : Le calcul d'une matrice de co-occurrence de niveaux de gris ou GLCM (Grey Level Co-occurrence Matrix) consiste à repérer dans une image le nombre d'occurrences de paires de niveaux de gris séparés par une distance d dans une direction définie par un vecteur de déplacement (dx, dy) .

Le calcul de la GLCM pour une image I de dimension $N \times M$ se formalise comme suit :

$$GLCM_{dx,dy}(i, j) = \sum_{n=1}^N \sum_{m=1}^M \begin{cases} 1, & \text{si } I(n, m) = i \text{ et } I(n + dx, m + dy) = j \\ 0, & \text{sinon} \end{cases}$$

Où i et j sont les niveaux de gris du pixel de référence et du pixel voisin respectivement, n et m correspondent aux coordonnées des pixels dans l'image I [31].

Indice de texture	Équation	Explication
Contrast	$f_1 = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$	Le contraste est une mesure de la variation d'intensité locale entre un pixel et son voisin sur toute l'image. Pour une image « constante » (sans variation), le contraste est nul.
Homogeneity	$f_2 = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$	L'homogénéité locale mesure la proximité de la distribution des éléments dans le GLCM avec la diagonale du GLCM. Pour un GLCM diagonal, l'homogénéité est de 1.
Energy	$f_3 = \sum_{i,j=0}^{N-1} (P_{i,j})^2$	L'ASM, également appelé énergie, est une mesure de l'homogénéité d'une image. Une scène homogène ne contiendra que quelques niveaux de gris mais des valeurs relativement élevées ; des valeurs plus élevées indiquent une uniformité de texture. Par conséquent, lorsque l'image est homogène, l'ASM aura des valeurs élevées.
Entropy	$f_4 = \sum_{i,j=0}^{N-1} P_{i,j} (-\ln P_{i,j})$	L'entropie mesure le caractère aléatoire de la texture de l'image (distribution d'intensité). Par conséquent, une image homogène entraînera une valeur d'entropie inférieure, tandis qu'une région hétérogène entraînera une valeur d'entropie plus élevée.
Dissimilarity	$f_5 = \sum_{i,j=0}^{N-1} P_{i,j} i-j $	La dissimilarité est une mesure de la distance entre des paires d'objets (pixels) dans la région d'intérêt. Par conséquent, une image homogène entraînera une dissimilarité plus faible.
Cluster shade	$f_6 = \sum_{i,j=0}^{N-1} [i+j-\mu_x(i)-\mu_y(j)]^3 P(i,j)$	La nuance du cluster est une mesure de l'asymétrie de la matrice. Lorsque la nuance du cluster est élevée, l'image est asymétrique.
Cluster prominence	$f_7 = \sum_{i,j=0}^{N-1} [i+j-\mu_x(i)-\mu_y(j)]^4 P(i,j)$	La proéminence du cluster est également une mesure de l'asymétrie. Lorsque la valeur de proéminence du cluster est élevée, l'image est moins symétrique.

TABLE 3.1 – Les caractéristiques de texture de GLCM utilisées dans notre étude [toutes les équations sont de [33] sauf l'équation de dissimilarité (ou FDM) est de [34]].

5 Classification : Par les forêts aléatoires (RF)

C'est la dernière étape dans un système de classification automatique des images médicales. Dans notre projet nous avons utilisé les forêts aléatoires ou RF (Random Forests) comme classifieur.

5.1 Classification

La classification automatique des images consiste à attribuer automatiquement une classe à une image à l'aide d'un système de classification. On retrouve ainsi la classification d'objets, de scènes, de textures, la reconnaissance de visages, d'empreintes digitale et de caractères.

Il existe deux principaux types d'apprentissage :

- l'apprentissage supervisé
- l'apprentissage non-supervisé

Dans *l'approche supervisée*, chaque image est associée à une étiquette qui décrit sa classe d'appartenance. Dans *l'approche non-supervisée* les données disponibles ne possèdent pas d'étiquettes. Dans notre travail on s'intéresse de l'approche supervisée [35].

5.1.1 Classification supervisée

L'objectif de la *classification supervisée* est principalement de définir des règles permettant de classer des objets dans des classes à partir de variables qualitatives ou quantitatives caractérisant ces objets.

On dispose au départ d'un échantillon dit *d'apprentissage* dont le classement est connu. Cet échantillon est utilisé pour l'apprentissage des règles de classement.

Il est nécessaire d'étudier la fiabilité de ces règles pour les comparer et les appliquer, évaluer les cas de sous apprentissage ou de sur apprentissage (complexité du modèle). On utilise souvent un deuxième échantillon indépendant, dit de *validation* ou de *test* [35].

5.2 Forêts aléatoires (ou Random Forests)

Quant au classification, nous proposons les *forêts aléatoires* ou *Random Forests (RF)*, qui ont été proposé par Leo Breiman en 2001 [36].

Définition

Soit $\{\hat{h}(\cdot, \Theta_1), \dots, \hat{h}(\cdot, \Theta_q)\}$ une collection de prédicteurs par arbre, où $(\Theta_1, \dots, \Theta_q)$ est une suite de variables aléatoires i.i.d., indépendante de l'échantillon d'apprentissage \mathcal{L}_n . Le prédicteur des forêts aléatoires est obtenu par agrégation de cette collection de prédicteurs [37].

RF est un algorithme d'apprentissage supervisé flexible qui fait bien partie de la famille des méthodes d'ensemble. Une forêt aléatoire est l'agrégation d'une collection d'arbres aléatoires généralement formés par la méthode de "bagging". L'idée générale de la méthode d'ensachage est qu'une combinaison de modèles d'apprentissage augmente le résultat global, il est largement utilisé, il peut être utilisé à la fois pour des tâches de classification comme de régression.

Principe

Le principe c'est de faire un apprentissage en parallèle sur de multiples arbres de décision construits aléatoirement et entraînés sur des sous-ensembles de données différents. Le nombre idéal d'arbres est très important il joue un rôle très efficace dans l'apprentissage.

Chaque arbre de la forêt aléatoire est entraîné sur un sous ensemble aléatoire de données selon le principe du bagging, en suivant le principe des « projections aléatoires ».

Un grand avantage de la forêt aléatoire est qu'elle peut être utilisée pour les problèmes de classification et de régression, qui constituent la majorité des systèmes d'apprentissage la machine actuels. Ci-dessous, vous pouvez voir à quoi ressemblerait une forêt aléatoire avec 600 arbres (figure 3.8).

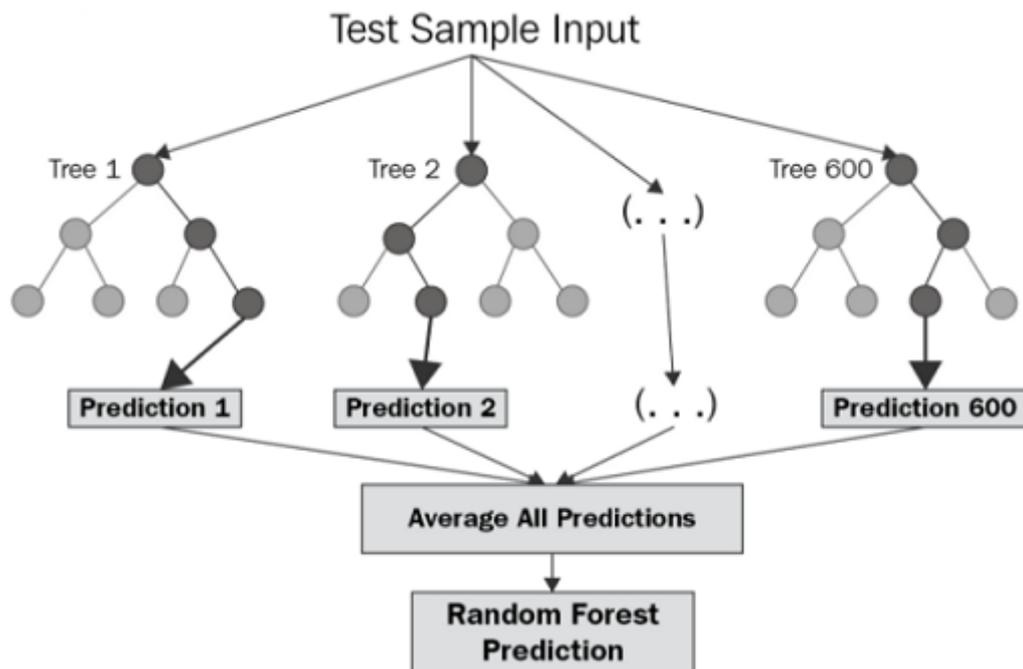


FIGURE 3.8 – Exemple d'une forêt aléatoire de 600 arbre [38].

6 Conclusion

Dans ce chapitre, on a expliqué les notions théoriques sur lesquelles on se base dans notre travail. On a discuter chaque étape de notre algorithme en détails nécessaire pour éclairer l'approche et la motivation derrière.

Pour le chapitre suivant, on va aborder les expérimentations et les résultats du programme obtenus tout en décrivant l'environnement et les outils de travail.

4

Chapitre 04 : Partie expérimentale

1 Introduction

Dans ce chapitre, nous allons entamer la partie expérimentale dans laquelle on va implémenter l'algorithme de classification des mammogrammes avec toutes ses étapes décrites en « *chapitre 3* ».

On va utilisé *Python* comme langage de programmation pour classifier un échantillon de la base de données *INbreast* en benign et maligne principalement par *RF* (*Random Forest*) et d'autres classifieurs pour comparer sa performance.

Enfin de compte, on va discuter les résultats et les avantages de cette étude.

2 La base de données : INbreast

La base de données *INbreast* est une base de données mammographiques, avec des images acquises dans un Centre du sein, situé dans un hôpital universitaire (*Hospital de São João, Centre du sein, Porto, Portugal*).

INbreast a un total de 115 cas (410 images), dont 90 cas de femmes ayant les deux seins (4 images par cas) et 25 cas de patientes ayant subi une mastectomie (2 images par cas). Plusieurs types de lésions (masses, calcifications, asymétries et distorsions) sont inclus (figure 4.1) [39].

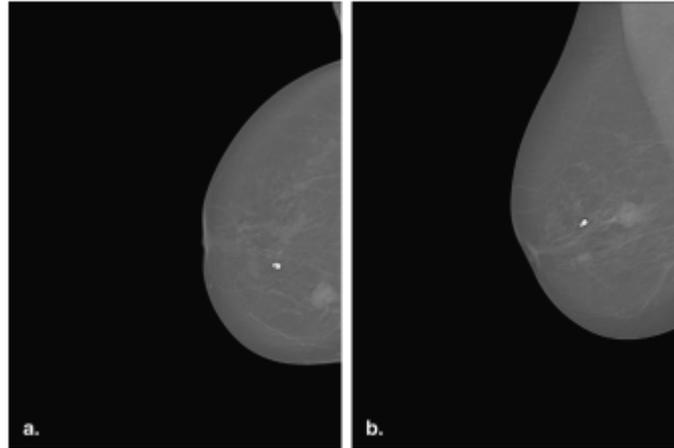


FIGURE 4.1 – Exemples des images dans INbreast [39].

2.1 Nos données : Protocole de test

Dans ce travail, on a utilisé 85 images de *INbreast* en format (.jpg) réparties en deux groupes, une base pour *l'apprentissage* du système qui inclut 73 images dont 10 sont labélisées comme malignes (ou cancéreuses) et une autre base pour *le test* qui inclut 12 nouvelles images dont 3 sont labélisées comme malignes (tableau 4.1).

	Bénignes	Malignes	Totale
Base d'apprentissage	63	10	73
Base de test	9	3	12

TABLE 4.1 – La répartition des données utilisées.

3 L'environnement de développement : Une brève description

3.1 Description du Hardware

Pour la machine sur laquelle on a exécuter notre programme, c'est un Laptop HP ProBook 450 G5 avec les capacités de calcul ci-dessous :

- **CPU** : Processeur Intel Core i5-8250U (8th Gen) (Quad-Core 1.6 GHz / 3.4 GHz Turbo - Cache 6 Mo)
- **RAM** : 8 Go de mémoire DDR4 2400 MHz (1x 8 Go - 2 slots)
- **Disque dur** : SSD M.2 SATA de 128 Go + disque dur de 1 To (5400 RPM)
- **GPU** : NVIDIA Geforce 930MX (2 GB de mémoire vidéo)

3.2 Description du Software

3.2.1 Python

On a utilisé le langage « *Python3* » (version 3 de Python sortie en version finale depuis décembre 2008) comme langage de programmation pour plusieurs avantages :

1. Un langage populaire et éminent en ce qui concerne la « *Data Science* » y compris *la classification* dont on est intéressé.
2. Sa syntaxe est conçue pour être lisible et directe ce qui met l'accent sur la lisibilité et réduit donc le temps et l'effort pour la *maintenance du code*.
3. Il offre une large gamme de *bibliothèques* dédiées pour la manipulation des images ce qui nous facilitera la phase de *pré-traitement*.

3.2.2 Anaconda : Jupyter notebook

On a utilisé *Jupyter notebook* (figure 4.2) de la distribution *Anaconda* comme *IDE* (*environnement de développement intégré*) pour créer et exécuter du code *Python*. *Jupyter notebook* est une application Web open source qui permet de créer et de partager des documents contenant du code en direct, des équations, des visualisations et du texte narratif. Il est structuré en cellules dont chaque cellule a sa propre contenu ce qui permet d'exécuter des morceaux de code indépendamment.

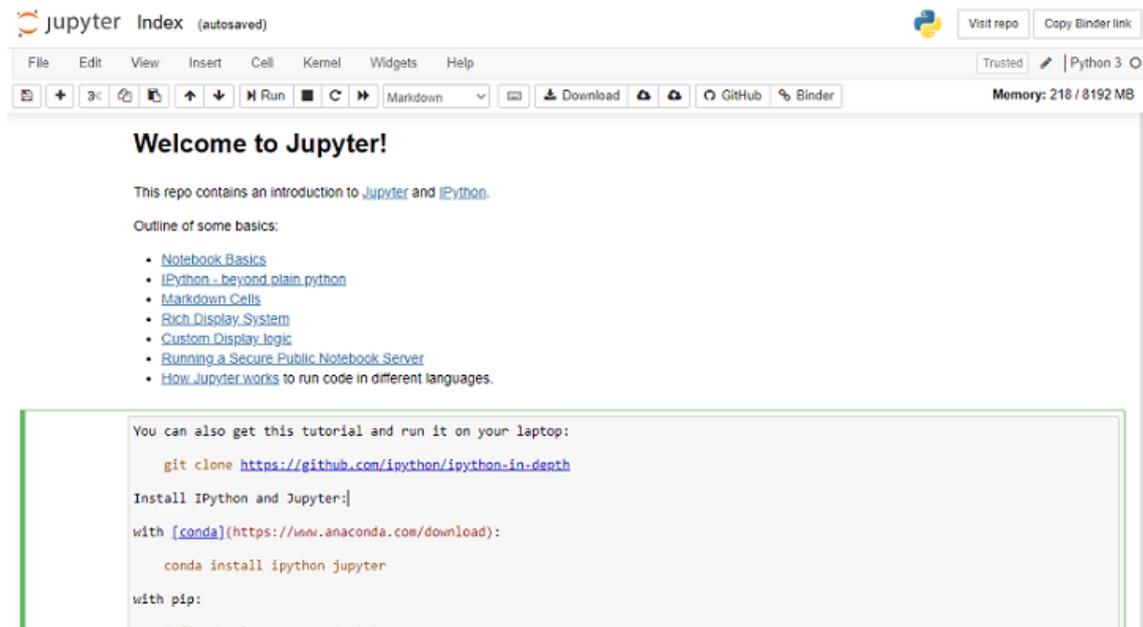


FIGURE 4.2 – Version demo en ligne de Jupyter notebook [40].

3.2.3 Bibliothèques utilisées

Au long de notre programme on a prit l'avantage de plusieurs bibliothèques de *Python* (figure 4.3) :

- **NumPy** : est une bibliothèque pour langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.
- **SciPy** : est un projet visant à unifier et fédérer un *ensemble de bibliothèques* Python à usage scientifique. Scipy utilise les tableaux et matrices du module NumPy.
- **OpenCV** : OpenCV (pour Open Computer Vision) est une bibliothèque graphique libre, initialement développée par Intel, spécialisée dans le traitement d'images en temps réel.
- **pandas** : est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles.
- **scikit-image** : (a.k.a. skimage) est une bibliothèque libre de traitement d'images en Python.
- **Matplotlib** : est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques. Elle peut être combinée avec les bibliothèques python de calcul scientifique NumPy et SciPy.



FIGURE 4.3 – Les logos des bibliothèques utilisées.

4 Implémentation et résultats

Dans cette partie, on va exposer l'implémentation de chaque étape de notre approche de classification des images de mammographie, chacune en discutant les résultats obtenus. La figure ci-dessous décrit notre système en diagramme (figure 4.4).

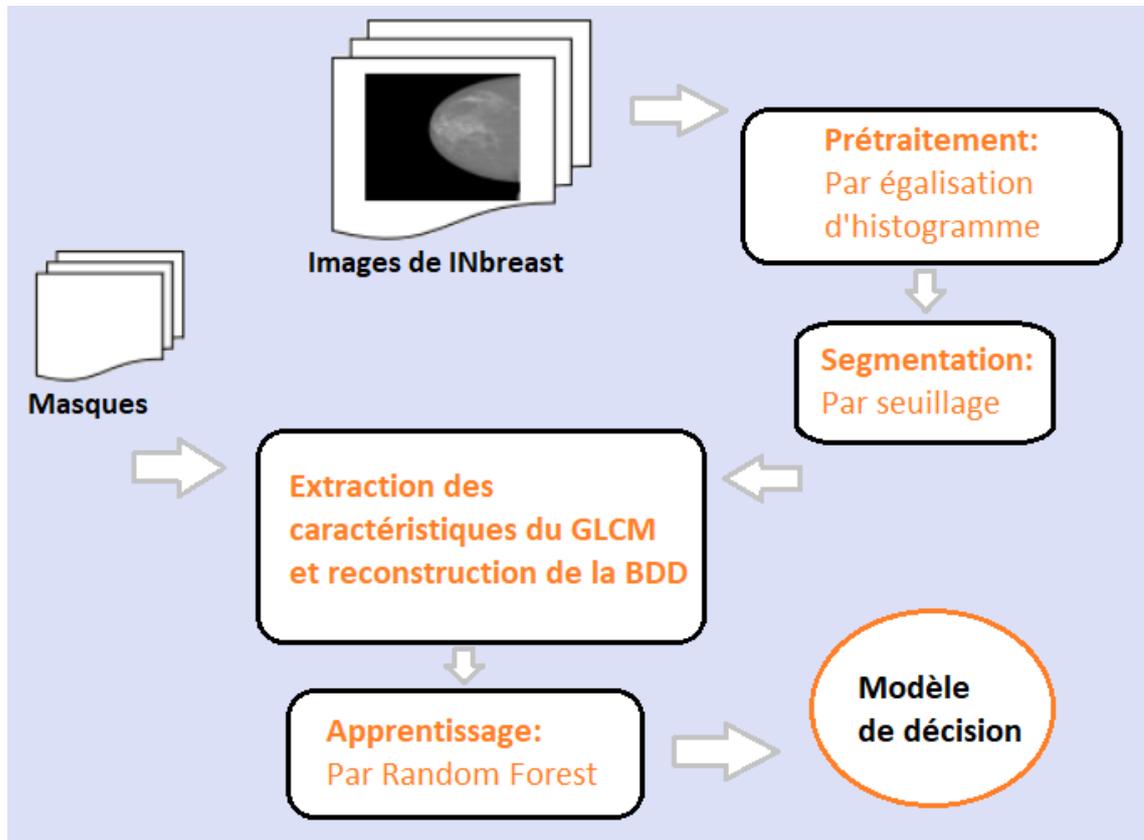


FIGURE 4.4 – Diagramme de notre système de classification des images mammographiques.

4.1 Prétraitements

Dans cette étape, on a redimensionné chaque image au taille de (512×512) , ensuite on a appliqué le filtre *median* et l'*égalisation adaptative d'histogramme* (CLAH) pour réduire le bruit impulsionnel et améliorer le contraste (Voire la Figure 4.5).

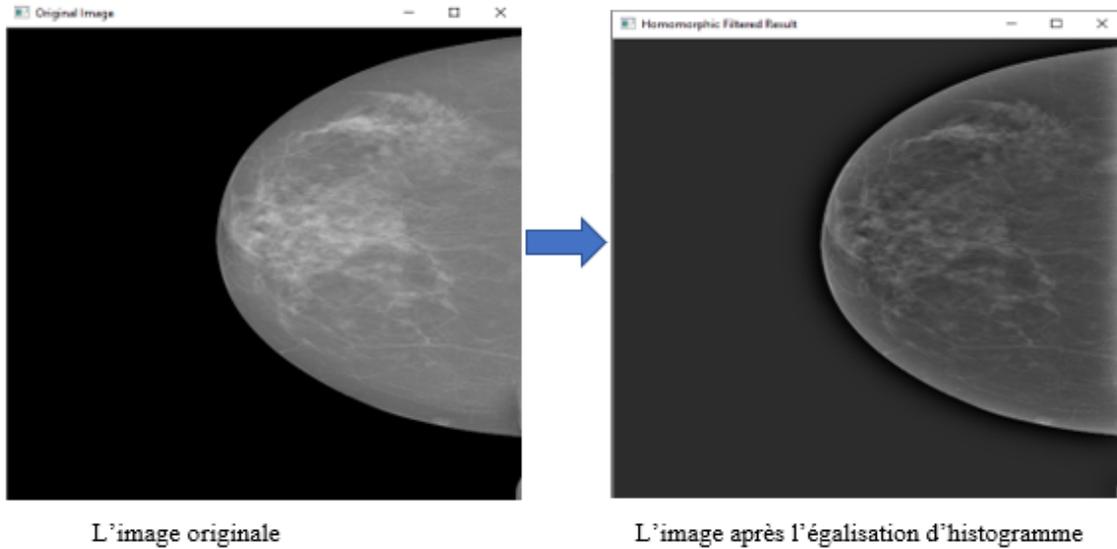


FIGURE 4.5 – Résultat du pré-traitement.

4.2 Segmentation

Sachant que la technique de seuillage global est rarement utilisée dans les mammographies car il y a de nombreuses structures dans les mammographies qui peuvent être une masse mais ne peuvent pas être révélées avec une seule seuil. Donc en général, les techniques de seuillage global ont une mauvaise performance par rapport aux techniques de seuillage local.

Dans cette étape, on a appliqué la technique de *seuillage adaptatif local* avec un seuil de « $S=190$ », le résultat de seuillage est présenté dans la figure 4.6. Ensuite on supprime les petites taches dans la zone de seuillage par les méthodes de reconstruction et morphologie (figure 4.7).

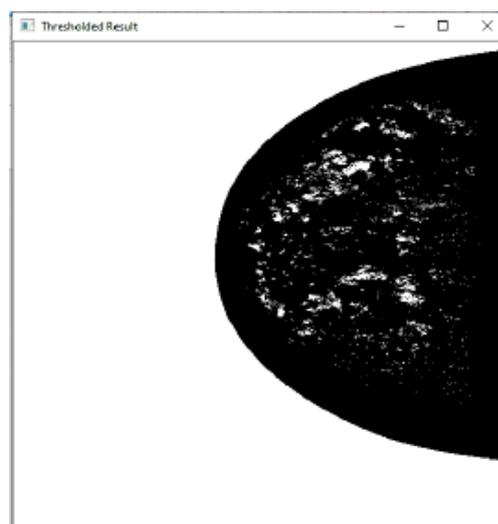


FIGURE 4.6 – L'image mammographie après la méthode de seuillage (binaire).

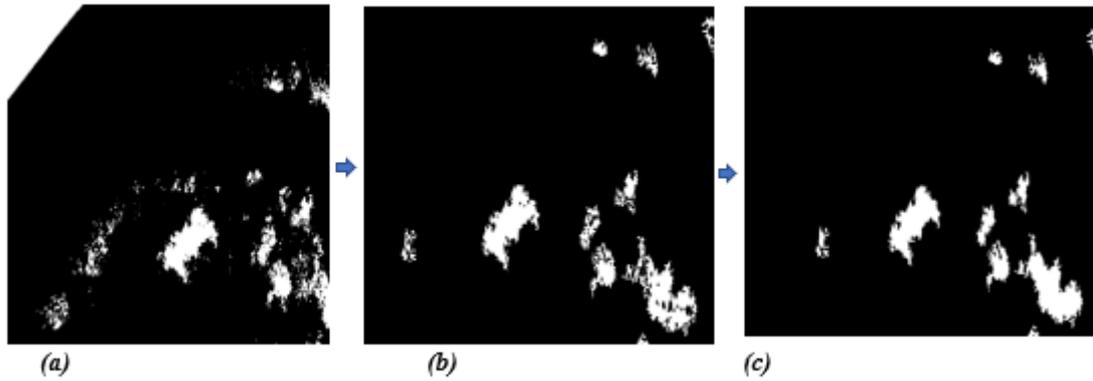


FIGURE 4.7 – La suppression des régions non unitives après le zoom d’image : a) seuillage , b) morphologie, et c) reconstruction

4.3 Caractérisation et apprentissage

Après la segmentation on fait extraire les 7 caractéristiques de texture préala-blement discutées dans le chapitre 3 (Contraste, homogénéité, énergie, entropie, dissimilarité, nuance du cluster, proéminence du cluster) pour chaque image avec lesquelles on construit notre base de données.

Ensuite On lance l’apprentissage du RF sur la base d’apprentissage qui contient les 7 valeurs de texture pour 73 images, après on procède au test sur la base de test qui contient les 7 valeurs de textures pour 12 images. Les résultats de classi-fication sont représentés dans le tableau 4.2.

Classifieur	Accuracy	Spécificité	Sensibilité
RF	0.66	1	0.42
SVM	1	1	1
RN	0.91	1	0.85

TABLE 4.2 – Les résultats des déférentes méthodes de classification.

4.4 Discussion des résultats

Pour approfondir l'analyse on a tracé la courbe de ROC pour le RF (figure 4.8), a partir de laquelle on peut observer que cette méthode de classification n'est pas vraiment capable de classer efficacement les malignes et bénignes dans une capacité qui peut être travaillée en toute confiance avec le taux de classification qui est relativement faible(0,66%).

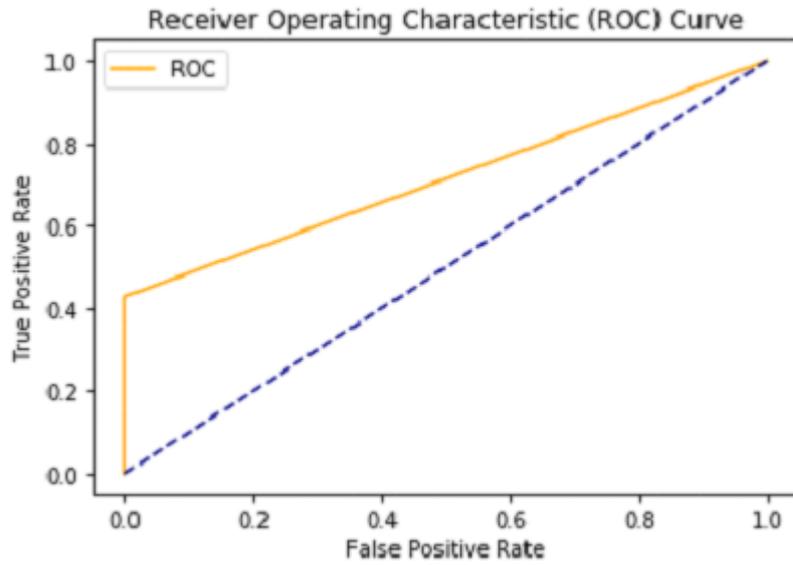


FIGURE 4.8 – Courbe de ROC pour le RF

Pour détecter à quelle partie se réside le problème (la partie de segmentation, d'apprentissage et caractérisation, ou dans la méthode de classification elle-même), on a testé d'autres classifieurs sur la même base de données qu'auparavant et on a comparé les résultats de classification dans la tableau 4.2.

A partir du tableau 4.2 nous observons que les classifieurs *SVM* ($ACC=1$) et *RN* ($ACC=0.91$) donnent des meilleurs Accuracy (ou taux de classification) par rapport au RF ($ACC=0.66$), mais basant sur les autres mesures la valeur de FP (Faux positive) est très élevée par rapport à RF, ce qui signifie que le problème ne se situe pas dans les autres étapes mais dans l'étape de classification et la nature de classificateur utilisé.

5 Conclusion

Dans ce chapitre, on a développé la partie expérimentale de ce travail en tout ce qui concerne les données (la base de données utilisé et sa répartition), l'outil de développement (le langage, les bibliothèques employés, et le compilateur du code) mais aussi chaque étape de l'implémentation de l'algorithme (pré-traitement, segmentation, caractérisation et l'apprentissage).

On a discuté les résultats obtenus figurés dans le tableau 4.2 pour chaque classifieur et la courbe de ROC 4.8 pour le RF, depuis lesquels on a constaté que le processus présenté donne de bons résultats avec le RN surtout mais il est besoin encore de raffinement et de développement pour achever de meilleurs résultats.

Conclusion et perspectives

Le cancer du sein devient le cancer le plus dangereux chez la femme, raison pour laquelle il est très célèbre, ce qui a incité les chercheurs à trouver différentes méthodes de DAO pour détecter la masse mammaire avant qu'elle ne se développe et devienne un cancer mortel.

Dans notre travail, on a discuté les étapes de la création d'un système d'aide au diagnostic pour aider les radiologues à la distinction et la classification des images mammographie en Malignes et Bénignes.

Notre approche se base sur différentes phases, premièrement on entame la phase de prétraitement des images de la base de données *INbreast* afin de récupérer la zone du sein avec un fond nettoyé, puis on utilise les approches de seuillage d'histogramme pour segmenter les différentes régions suspectées dans l'image. Par la suite, on procède au phase d'extraction des caractéristiques en utilisant les descripteurs de type GLCM afin d'aboutir au classification avec succès.

A la fin, un classifieur est entraîné sur les caractéristiques extraites, ce classifieur est capable de classer les masses mammaires en maligne ou bénigne. Les caractéristiques du GLCM sont très efficaces pour l'apprentissage du classifieur qui a été prouvée par deux classifieurs : SVM et RN (Réseaux de neurones), ces derniers donnent de meilleurs résultats de classification en termes de *taux de classification*, de *spécificité* et de *sensibilité*.

Dans la perspective de mettre en place un certain nombre d'études pour mieux développer cette approche dans le futur, tout en proposant d'utiliser d'autre technique de segmentation plus performante comme la *LPE* ou le *Deep Learning*.

Aussi, nous proposons d'enrichir la phase de caractérisation par d'autres descripteurs capable d'améliorer les performances de la classification des masses mammaires.

Bibliographie

- [1] J. E. E. de OLIVEIRA, A. de ALBUQUERQUE ARAÚJO et T. M. DESERNO, « Content-based image retrieval applied to BI-RADS tissue classification in screening mammography, » *World journal of radiology*, t. 3, n° 1, p. 24, 2011.
- [2] *Breast cancer now most common form of cancer : WHO taking action*. adresse : <https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action> (visité le 03/02/2021).
- [3] E. KOZEGAR, M. SORYANI, B. MINAEI, I. DOMINGUES et al., « Assessment of a novel mass detection algorithm in mammograms, » *Journal of cancer research and therapeutics*, t. 9, n° 4, p. 592, 2013.
- [4] C. DUVALLET, « Des systèmes d'aide à la décision temps réel et distribués : modélisation par agents, » thèse de doct., Université du Havre, 2001.
- [5] *Anatomie du sein*. adresse : <http://www.depistagesein.ca/anatomie-du-sein> (visité le 14/06/2020).
- [6] *L'anatomie du sein – pathologies bénignes et malignes du sein*. adresse : <https://www.docteur-eric-sebban.fr/cancer-du-sein/diagnostic-cancer-sein/anatomie-et-pathologies-du-sein/> (visité le 26/06/2020).
- [7] *Qu'est-ce que le cancer du sein?* Adresse : <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/> (visité le 02/07/2020).
- [8] *Quels sont les différents types de tumeurs?* Adresse : <https://www.beljanski.info/quels-sont-les-differents-types-de-tumeurs/> (visité le 03/07/2020).
- [9] *Tumeur au sein*. adresse : <https://cancer-du-sein.ooreka.fr/astuce/voir/621915/tumeur-au-sein> (visité le 05/07/2020).
- [10] P. P. FOURET, *Anatomie pathologique : Chapitre 4 - Pathologie tumorale*. adresse : <http://www.chups.jussieu.fr/polys/anapath/Cours/POLY.Chp.4.html> (visité le 10/07/2020).
- [11] *Qu'est-ce qu'une mammographie?* Adresse : <https://www.biron.com/fr/imagerie-medicale/mammographie/radiographie-des-seins/> (visité le 12/07/2020).
- [12] I. CHEIKHROUHOU, « Description et classification des masses mammaires pour le diagnostic du cancer du sein, » thèse de doct., 2012.

- [13] E. A. SICKLES, C. J. D'ORSI, L. W. BASSETT, C. M. APPLETON, W. A. BERG, E. S. BURNSIDE et al., « ACR BI-RADS® Atlas, Breast imaging reporting and data system, » *Reston, VA : American College of Radiology*, p. 39-48, 2013.
- [14] J. N. WOLFE, « Breast patterns as an index of risk for developing breast cancer, » *American Journal of Roentgenology*, t. 126, n° 6, p. 1130-1137, 1976.
- [15] A. LE TREUT et M.-H. DILHUYDY, *Mammographie : guide d'interprétation*. 1988.
- [16] J. SUCKLING, D. DANCE, E. MOSKOVIC, D. LEWIS et S. BLACKER, « Segmentation of mammograms using multiple linked self-organizing neural networks, » *Medical Physics*, t. 22, n° 2, p. 145-152, 1995.
- [17] S. SUKUMARAN, « A Review : Breast Mass Classification in Mammogram Images, » *International Journal of Engineering Research*, t. 3, n° 4, 2014.
- [18] S. I. M. ALI et N. H. SALMAN, « Automatic Segmentation of Breast Mammograms Using Hybrid Density Slicing and k-mean Adaptive Methods, » *Journal of Engineering and Applied Sciences*, t. 14, n° 15, p. 5044-5050, 2019.
- [19] R. DUBEY, M. HANMANDLU et S. GUPTA, « Segmentation and characterization of masses in the digital mammograms, » 2010.
- [20] H. ZHANG, S. W. FOO, S. M. KRISHNAN et C. H. THNG, « Automated breast masses segmentation in digitized mammograms, » in *IEEE International Workshop on Biomedical Circuits and Systems, 2004.*, IEEE, 2004, S2-2.
- [21] S. M. BADAWY, A. A. HEFNAWY, H. E. ZIDAN et M. T. GADALLAH, « Breast cancer detection with mammogram segmentation : A qualitative study, » *International Journal of Advanced Computer Science and Application*, t. 8, n° 10, 2017.
- [22] A. MAKANDAR et B. HALALLI, « Threshold based segmentation technique for mass detection in mammography, » *J Comput*, t. 11, n° 6, p. 472-478, 2016.
- [23] M. A. MAZUROWSKI, J. Y. LO, B. P. HARRAWOOD et G. D. TOURASSI, « Mutual information-based template matching scheme for detection of breast masses : From mammography to digital breast tomosynthesis, » *Journal of biomedical informatics*, t. 44, n° 5, p. 815-823, 2011.
- [24] M. M. ELTOUKHY et I. FAYE, « An adaptive threshold method for mass detection in mammographic images, » in *2013 IEEE International Conference on Signal and Image Processing Applications*, IEEE, 2013, p. 374-378.
- [25] J. s. JEBAMONY et D. JACOB, « Classification of Benign and Malignant Breast Masses on Mammograms for Large Datasets using Core Vector Machines, » *Current Medical Imaging*, t. 16, n° 6, p. 703-710, 2020.
- [26] H. A. KHAN, A. AL HELAL, K. I. AHMED et R. MOSTAFA, « Abnormal mass classification in breast mammography using rotation invariant LBP, » in *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, IEEE, 2016, p. 1-5.
- [27] R. F. dos SANTOS TEIXEIRA, « Automatic analysis of mammography images : classification of breast density, » 2013.

- [28] M. BERGOUNIOUX, *Introduction au traitement mathématique des images-méthodes déterministes*. Springer, 2015, t. 76.
- [29] M. E. BRIKI, « Medical Images Classification based on deep features extraction exploiting transfer learning, » mém. de mast., Abou Bekr Belkaid university of Tlemcen, nov. 2020.
- [30] M. SONKA, V. HLAVAC et R. BOYLE, *Image processing, analysis, and machine vision*, Fourth edition. Cengage Learning, 2014.
- [31] N. E. H. SAYAH et S. DJELTI, « ETUDE DES METHODES D'ANALYSE DE LA TEXTURE DES IMAGES MEDICALES ARM, » mém. de mast., Abou Bekr Belkaid university of Tlemcen, sept. 2017.
- [32] R. M. HARALICK, K. SHANMUGAM et I. H. DINSTEN, « Textural features for image classification, » *IEEE Transactions on systems, man, and cybernetics*, n° 6, p. 610-621, 1973.
- [33] X. YANG, S. TRIDANDAPANI, J. J. BEITLER, D. S. YU, E. J. YOSHIDA, W. J. CURRAN et T. LIU, « Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy : An in vivo study of late toxicity, » *Medical physics*, t. 39, n° 9, p. 5732-5739, 2012.
- [34] R. XU, X. ZHAO, X. LI, C. KWAN et C. CHANG, « Target detection with improved image texture feature coding method and support vector machine, » *International Journal of Intelligent Technology*, t. 1, n° 1, p. 47-56, 2006.
- [35] M. Z. MOKRI, « Classification des images avec les réseaux de neurones convolutionnels, » mém. de mast., Abou Bekr Belkaid university of Tlemcen, juill. 2017.
- [36] L. BREIMAN, « Random forests, » *Machine learning*, t. 45, n° 1, p. 5-32, 2001.
- [37] R. GENUER, « Forêts aléatoires : aspects théoriques, sélection de variables et applications, » thèse de doct., 2010.
- [38] C. BAKSHI, *Random Forest Regression*. adresse : <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84> (visité le 15/08/2020).
- [39] I. C. MOREIRA, I. AMARAL, I. DOMINGUES, A. CARDOSO, M. J. CARDOSO et J. S. CARDOSO, « Inbreast : toward a full-field digital mammographic database, » *Academic radiology*, t. 19, n° 2, p. 236-248, 2012.
- [40] *Welcome to Jupyter!* Adresse : <https://notebooks.gesis.org/binder/jupyter/user/ipython-ipython-in-depth-ce5tvk8/notebooks/binder/Index.ipynb> (visité le 03/09/2020).
- [41] J.-L. ROSE, « Croissance de région variationnelle et contraintes géométriques tridimensionnelles pour la segmentation d'image, » thèse de doct., Lyon, INSA, 2008.
- [42] B. SÉNÉCAIL et B. MENANTEAU, « Anatomical basis for tomographic exploration of the suprarenal glands, » *Anatomia Clinica*, t. 4, n° 2, p. 93-100, 1982.
- [43] E. J. SMITH et J. H. WOLFE, « Observations of interaction regions and corotating shocks between one and five AU : Pioneers 10 and 11, » *Geophysical Research Letters*, t. 3, n° 3, p. 137-140, 1976.

-
- [44] A. T. STAVROS, D. THICKMAN, C. L. RAPP, M. A. DENNIS, S. H. PARKER et G. A. SISNEY, « Solid breast nodules : use of sonography to distinguish between benign and malignant lesions. », *Radiology*, t. 196, n° 1, p. 123-134, 1995.
- [45] N. D. HAMADOUCHE MEROUA, « Détection et classification des tumeurs dans les tissus mammaires par la machine à vecteur de support(SVM), » mém. de mast., 2019.
- [46] B. BEN ALI TENFIR, « Segmentation des images mammographiques en vue de la détection et la caractérisation Des masses mammaires, » thèse de doct.
- [47] C. DESIR, « Classification automatique d'images, application à l'imagerie du poumon profond, » thèse de doct., 2013.
- [48] A. KEFALI, T. SARI et M. SELLAMI, « Evaluation of several binarization techniques for old Arabic documents images, » in *The First International Symposium on Modeling and Implementing Complex Systems MISC*, t. 1, 2010, p. 88-99.
- [49] I. KESSOUR et T. IMANE, « Simulation des contours actifs par les colonies de fourmis, » *Pour l'obtention du diplôme d'Ingénieur d'Etat en Informatique*, 2011.
- [50] M. HALL-BEYER, « GLCM texture : a tutorial v. 3.0 March 2017, » 2017.