



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 البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

présenté par

Djebari Mohamed

&

Kerkache Akram

pour l'obtention du diplôme de Master en Électronique option 'vision et robotique'

Thème

Détection et classification connexionniste de pathologies mammographiques

Proposé par : Reguieg F.Zohra & Benblidia Nadjia

Année Universitaire 2012-2013



Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience, d'accomplir ce travail.

En second lieu, nous tenons à remercier notre promotrice Mlle **Reguieg**, pour son encadrement, sa disponibilité, ses précieux conseils, son encouragement et son aide durant, toute la période du travail.

Nous remercions aussi, notre co-promotrice Mlle **Benblidia** pour sa disponibilité et son aide qui nous a été très utile.

Nos vifs remerciements vont également aux membres du jury, pour l'intérêt qu'ils ont porté à notre préliminaire à la recherche, en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

Enfin, nous tenons également à remercier toutes les personnes, qui ont participé de près ou de loin à la réalisation de ce travail que ce soit en nous aidant, nous encourageant ou nous montrant simplement la voie.



Dédicaces

Louage à dieu, Seigneur du monde

C'est toi que nous adorons et de toi, que nous implorons secours. Que la paix et la bénédiction soit sur son dernier envoyé.

A mes chers parents qui ont été présents, à chaque instant, leur Irremplaçable et inconditionnel soutien m'a permis d'écarter les doutes, de soigner les blessures et partager mes joies.

*A mon frère **wail** et mes deux belles sœurs*

*A mon beau frère **Rabah** et ces enfant **louai** et **rafik***

A toute ma famille, mes tantes et mes oncles

A mes amis: Samir, Mohamed, Hocine, Azzedine, afaf

A tous mes amis sans exception

A ma belle fiancée

Kerkache Akram



Dédicaces

Je dédie travail à ceux qui m'ont appris ce qu'est la valeur du travail, à ma mère celle qui m'a toujours soutenu dans tous mes projets; à mon père qui m'a tout donné, pour que je réussisse dans la vie; merci car sans vous, je ne serais point où je suis arrivé aujourd'hui.

*Je n'oublie pas de dédier mon modeste travail, à mes trois frères, **Mehdi, Merouane** et **Yacine**, à mes belles sœurs, et spécialement à ma nièce, **Louisa-Kawther** et, mon neveu **Mohamed-Hicheme**.*

Je dédie aussi ce travail à mes amis, tous sans exception, à mes camarades de la promotion, à mes compagnons de la résidence universitaire, à mes vieux amis et nouveaux, tous on été là à un moment ou à un autre, pour me soutenir et m'encourager, je savais qu'avec eux, rien ne pouvait m'arriver, car avec eux, j'étais sur des épaules de géant!

DJEBARI Mohamed



ملخص:

هذا العمل يقدم طريقة لاكتشاف الامراض في التصوير الرقمي للثدي ، مع التركيز على تصنيف الارتباطية. المعالجة المستخدمة تركز على الكشف عن تكتلات الكتل و الحالات الصحية ، لهذا الغرض استعملنا تجزئة تعاونية بين تقنية تحليل الرسم البياني و النماذج المشوهة ، و هذا لاستخراج المناطق المثيرة للاهتمام بتشخيصها من خلال الصفات المورفولوجية و التكوينية (النسيجية). استخدم هذا الوصف في مدخل مصنف الشبكة العصبية بعد هذا لتحديد الخمسة فئات و لمساعدتنا في اتخاذ القرار استعملنا مستقبل متعدد الطبقات بالإضافة الى الخوارزميات الجينية. كلمات دالة : صورة تصوير الثدي ، نماذج للتشوه ، تجزئة تعاونية ، النسيج و التشكل ، تصنيف الارتباطية ، الخوارزميات الجينية

Résumé : ce travail présente une méthode de détection de pathologies en mammographie numérique, en se focalisant sur une classification connexionniste. Le traitement utilisé, se concentre sur la détection des calcifications, des masses et des cas sains. A cet effet, une segmentation coopérative entre la technique par analyse d'histogramme et les modèles déformables, est réalisée, afin d'extraire les régions d'intérêt pour les caractériser grâce à des paramètres morphologiques et texturaux. Cette description est utilisée dans l'entrée d'un réseau de neurones classificateur, le perceptron multicouche associé aux algorithmes génétiques, pour l'identification de cinq classes, en vue d'une aide à la décision.

Mots clés : image mammographique, modèles déformables, segmentation coopérative, texture et morphologie, classification connexionniste, algorithmes génétiques, aide à l'interprétation.

Abstract: : This work presents a method for detection of pathologies in digital mammography, focusing on a connectionist classification. The treatment used, focuses on the detection of calcifications, masses and healthy cases. For this purpose, a cooperative segmentation between the technique by histogram analysis and distortable patterns is performed to extract the regions of interest as to characterize them through morphological and textural parameters. This description is used in the input of a neural network classifier, the multilayer perceptron model combined with genetic algorithms for the identification of five classes, for a decision support.

Keywords: mammography image, distortable patterns, cooperative segmentation, texture and morphology, connectionist classification, genetic algorithms, interpretation support.





Listes des acronymes et abréviations

DA : distorsions architecturales

BI-RADS: Breast Imaging Reporting and Data System

CA: contour actif

CHF : chapeau haut forme

RN : réseau de neurones

SVM : machines à support de vecteur (support vector machine)

ENE : énergie

CST : contraste

ENT : entropie

VAR : variance

COR: corrélation

IDM: moment inverse

Comp: compacité

AG: algorithme génétique

Pmc: perceptron multi-couche

SVNN: support Vector neural network

ROC: *Receiver* Operating Characteristic Curve



Table des matières

Introduction général	1
Chapitre 1 contexte mammographie et cancer du sein	3
1.1) anatomie de sein.....	4
1.2) Introduction au Cancer de sien.....	4
1.2.1) symptômes de cancer de seins	6
1.2.2) facteur de risque.....	6
a) facteur internes.....	6
b) facteur externes	7
1.2.3) catégories de cancer du sein.....	7
a) Cancer non invasif	7
b) Cancer invasif ou infiltrat.....	7
1.3) imagerie mammographie	9
1.3.1) Principe de la mammographie	9
a) numérisation indirecte	10
b) La mammographie numérique	10
1.3.2) Type de l'examen	11
a) L'examen de dépistage	11
b) L'examen diagnostique.....	11
1.4) Pathologies mammaires	12
1.4.1) Kyste.....	12
1.4.2) Masses	12
a) Distorsions architecturales	13
b) Asymétrie de densité	13
c) Lésions spéculées	14
d) Classification des masses mammaire	15
1.4.3) Calcifications.....	15
1.5) Caractéristiques radiologiques.....	16
1.5.1) Spécificité et Sensibilité.....	16
1.5.2) Faux positifs et faux négatifs.....	17
1.5.3) Densité mammaire.....	18
1.5.4) Classification de la densité mammaire.....	18
1.6) Conclusion.....	19
Chapitre 2 Techniques de segmentation en imagerie et méthodologie adoptée	
2.1) Introduction.....	20
2.2) Principe de la segmentation.....	21
2.3) Différentes méthodes de segmentation.....	23
2.3.1) Approche région.....	23
a) Segmentation par division de régions.....	24
b) Segmentation par fusion de régions.....	24
c) Segmentation par division fusion.....	25
d) Segmentation par croissance de région	26
2.3.2) Segmentation par les K-moyennes.....	28



2.3.3) Segmentation par analyse d'histogramme.....	29
a) Les méthodes de seuillage global	29
b) Les méthodes de seuillage local	29
2.3.4) Segmentation par contour	30
a) Segmentation par contours actifs	31
b) Segmentation par les ensembles de niveaux (levels sets)	34
2.3.5) Approche coopérative	36
a) Coopération séquentielle.....	36
b) Coopération des résultats	37
c) Coopération mutuelle	37
2.4) Introduction à la morphologie mathématique.....	38
2.4.1) Operateur de base et transformation premières de la morphologie Mathématique.....	38
2.4.2) Le gradient morphologique	42
2.4.3) Transformation chapeau haut de forme.....	42
2.5) Aperçu superficiel des travaux en imagerie mammographique.....	43
2.6) Méthodologie adoptée de détection des masses et calcifications.....	44
2.6.1) Cas des calcifications.....	44
2.6.2) Cas des masses.....	46
2.7 Conclusion	47

Chapitre 3: Caractérisation d'images mammographiques et classification connexionniste

3.1 Introduction	48
3.2 Analyse de texture.....	49
3.2.1 Définition de la texture	49
3.2.2 Types de texture	50
a. Texture périodique.....	50
b. Textures aléatoire	50
3.2.3 Introduction aux méthodes d'extraction des caractéristiques texturales.....	51
a. Méthodes structurales	51
b. Méthodes statistiques	51
c. Méthodes fractales	51
d. Méthodes fréquentielles	52
3.1.4 Matrices de concurrence	52
a. Energie.....	53
b. Contraste.....	53
c. Entropie.....	53
d. Variance.....	54
e. Corrélation.....	54
f. Moment inverse.....	54
3.3 Analyse de la forme.....	54
3.3 .1 Compacité.....	55
3.3.2 Excentricité.....	55
3.3.3 Convexité et enveloppe convexe	55
3.3.4 Moments invariants de Hu	56



3.4 Introduction à la classification d'image	57
3.4.1 Apprentissage.....	58
3.4.2 Apprentissage supervisé.....	59
3.5 Réseaux de Neurones.....	59
3.5.1 Neurones biologiques.....	60
3.5.2 Neurones artificiels	60
3.5.3 Principe de fonctionnement.....	61
3.5.4 Calcul des poids synaptiques.....	61
3.5.5 Les grands types de réseaux.....	61
a. Réseaux de Hopfield.....	61
b. Réseaux de Kohonen.....	62
c. Perceptrons	62
1. Perceptron simple	62
2. Perceptron multicouche (PMC).....	62
3.5.6 Fonction d'activation.....	64
3.5.7 Avantages et inconvénients.....	64
3.6 Algorithme génétique.....	65
3.6.1 Définition	65
3.6.2 Operateur génétique	65
a. sélection.....	65
b. croisement.....	66
c. mutation.....	66
3.6.3 Principe de fonctionnement.....	66
3.6.3 Avantages d'algorithme génétique	67
3.7 Classification en SVNN	68
3.7.1 Création de la base d'apprentissage.....	69
3.8 Evaluation du résultat.....	72
3.8.1 Courbe roc.....	72
3.8.2 matrice de confusion.....	73
3.8.3 Distance de mahalanobis.....	73
3.9 Conclusion	74
Chapitre 4.....	75
4.1. Environnement de travail.....	75
4.1.1 Matériel utilisé.....	75
4.1.2 Qu'est ce que Matlab?.....	76
4.1.3 Débuter en Matlab.....	76
4.1.4 Présentation du 'gui'	77
4.2 Présentation du système développé.....	79
4.2.1 Images utilisées.....	81
4.2.2 Interface du système.....	81



4.3 Résultats et interprétation.....	83
4.3.1 Détection des masses	83
4.3.3 Détection de calcifications.....	85
4.3.4 Extraction de caractéristiques.....	87
4.3.5 Classification.....	91
4.3.6 Evaluation des résultats.....	92
4.4 Conclusion.....	103
Conclusion générale et perspectives.....	104
Bibliographie.....	106



Liste des figures

Chapitre 1

Figure 1.1 : structure du sein	4
Figure 1.2 : sein cancérisé.....	5
Figure 1.3 : différent type de cancer.....	8
Figure 1.4 : mammographe numérique.....	9
Figure 1.5 : Géométrie d'acquisition d'une mammographie (vue de face et vue de profil).....	10
Figure 1.6 : clichés associés.....	11
Figure 1.7 : Exemple d'une masse en mammographie.....	13
Figure 1.8 : exemple d'une distorsion architecturale.....	13
Figure 1.9 : exemple d'asymétrie : une région suspecte dans le sein droit.....	14
Figure 1.10 : exemple de masses spéculées	14
Figure 1.11 : formes schématiques des masses.....	15
Figure 1.12 : contours schématiques des masses.....	15
Figure 1.13 : différentes formes des microcalcifications.....	16
Figure 1.14 : exemple de microcalcifications.....	16
Figure 1.15 : différents types de la densité mammaire.....	19

Chapitre 2

Figure 2.1 : arbre quaternaire sur une image 8x8	24
Figure 2.2 : image originale	25
Figure 2.3 : graphe d'adjacence.....	25
Figure 2.4 : Illustration du concept pyramidal de l'image pour la segmentation par fusion et Division.....	26
Figure 2.5 : application de l'algorithme croissance de région sur les images mammographiques.....	27
Figure 2.6 : application de l'algorithme k-means sur les images mammographiques.....	28
Figure 2.7 : Exemples de contours	31
Figure 2.8 : illustrations des grandes étapes du déroulement d'un contour actif.....	34
Figure 2.9 : schéma de la coopération séquentielle.....	37
Figure 2.10 : synoptique de la coopération des résultats.....	37
Figure 2.11 : synoptique de la coopération mutuelle.....	38
Figure 2.12 : érosion d'une image mammographique.....	39
Figure 2.13 : dilatation d'une image mammographique.....	40
Figure 2.14 : ouverture d'une image mammographique par un élément structurant circulaire.....	40
Figure 2.15 : fermeture d'une image mammographique par un élément structurant circulaire.....	41
Figure 2.16 : Gradient de Beucher d'une fonction f.....	42
Figure 2.17 : chapeau haut de forme, blanc WTH d'une fonction numérique f par un Élément structurant B.....	43
Figure 2.18 : chapeau haut de forme noir BTH d'une fonction numérique f par un élément Structurant B [19].....	43
Figure 2.19 : Détection des micro calcifications par chapeau haut de forme.....	45
Figure 2.20 : système de détection des calcifications.....	45
Figure 2.21 : système de détection des masses.....	46



Chapitre3

Figure 3.1 : Exemples de textures périodiques.....	50
Figure 3.2 : Exemples de textures aléatoires.....	50
Figure 3.3: pixel x selon quatre directions.....	52
Figure 3.4 : enveloppe convexe.....	55
Figure 3.5: modules d'un système d'apprentissage.....	59
Figure 3. 6 : neurone biologique.....	60
Figure 3.7 : représentation schématique d'un neurone artificiel.....	60
Figure 3.8 : perceptron simple.....	62
Figure 3.9: perceptron multicouche	63
Figure 3.10 : principe de fonctionnement AG.....	67
Figure 3.11 Représentation de la fonction sigmoïde	69
Figure 3.12 étapes d'apprentissage du réseau PMC	71
Figure 3.13: l'approche de la classification et divisions des classes.....	72

Chapitre4

Figure 4.1: fenêtre principale du guide.....	77
Figure 4.2 : Property Inspector.....	78
Figure 4.3 : système d'aide à la décision mammographique.....	80
Figure 4.4: image mammographiques de la base MIAS.....	81
Figure 4.5 : interface de l'application.....	82
Figure 4.6 : détection d'une masse maligne par les contours actifs.....	84
Figure 4.7 : détection d'une masse maligne par analyse d'histogramme.....	84
Figure 4.8 : détection d'une masse maligne par coopération.....	84
Figure 4.9 : détection de la forme des calcifications par les contours actifs (mdb209, nombre d'itérations1500)	86
Figure 4.10 : détection des calcifications par analyse d'histogramme (seuil= 150).....	86
Figure 4.12 : détection de calcification par coopération et isolation par chapeau haut de forme....	86
Figure 4.13: influence de la surface sur les masses.....	89
Figure 4.14: influence du contraste sur les masses.....	89
Figure 4.15 : influence de la surface sur les calcifications.....	90
Figure 4.16: influence de la compacité sur les calcifications.....	90
Figure 4.17: entraînement du réseau de neurones par les algorithmes génétiques.....	92
Figure 4.18 : matrice de confusion pour la catégorie des tests.....	93
Figure 4.19 : classification d'une masse maligne (mdb274)	94
Figure 4.20 : classification d'une masse maligne (mdb170)	94
Figure 4.21 : classification d'une masse maligne (mdb155)	95
Figure 4.21.a : résultat de l'analyse de l'histogramme	95
Figure 4.21.b : résultat du contour actif.....	95
Figure 4.22 : classification d'une masse bénigne (mdb097).....	96
Figure 4.23 : classification d'une masse bénigne (mdb199)	96
Figure 4.24 : classification d'une masse bénigne (mdb204)	97
Figure 4.24.a : résultat de l'analyse de l'histogramme.....	97
Figure 4.24.b : résultat du contour actif.....	97
Figure 4.25: classification de calcifications bénignes (mdb 218)	98
Figure 4.26: classification de calcifications bénignes (mdb219)	98



- Figure 4.26.a** : résultat de l'analyse de l'histogramme 99
- Figure 4.26.b** : résultat du contour actif..... 99
- Figure 4.27**: classification de calcifications malignes (mdb226) 99
- Figure 4.28**: classification de calcifications malignes (mdb223) 100
- Figure 4.29**: classification de calcifications malignes (mdb209) 100
- Figure 4.29.a** : résultat de l'analyse de l'histogramme..... 101
- Figure 4.29.b** : résultat du contour actif..... 101
- Figure 4.30**: classification d'un cas sain (mdb027)..... 101
- Figure 4.31**: classification d'un cas sain (mdb055)..... 102
- Figure 4.32**: classification d'un cas sain (mdb087) 102



Liste des tableaux

Chapitre 1

Tableau 1.1 : critères de distinction entre tumeurs bénignes et malignes.....5

Chapitre4

Tableau 4. 1 : paramètres de segmentation de masses malignes.....83

Tableau 4. 2 : paramètres de segmentation de masses bénignes.....83

Tableau 4.3 : paramètres de segmentation de calcifications malignes.....85

Tableau 4.4 : paramètres de segmentation de calcifications bénignes.....85

Tableau 4.5: paramètres des images de la classe des masses malignes.....87

Tableau 4.6: paramètres des images de la classe des masses bénignes87

Tableau 4.7: paramètres des images de la classe des calcifications malignes.....88

Tableau 4. 8: paramètres des images de la classe des calcifications bénignes.....88

Tableau 4.9: paramètres des images de la classe des cas sains.....88



Introduction générale

Le cancer du sein vient en tête des cancers répandus, en Algérie avec 10.000 cas enregistrés par an et, une moyenne d'âge de 40 dont 95% des cas qui se présentent à un stade avancé de la pathologie, ce qui empêche de prendre les bonnes options pour contrer cette maladie. C'est pour cela que l'étape du diagnostic est une étape clef dans la lutte contre le cancer du sein comme pour toute autre pathologie. C'est dans ce cadre, que s'inscrit le travail que nous allons présenter dans notre mémoire, intitulé « Détection et classification connexionniste de pathologies mammographiques »

Nous nous concentrerons sur les images qui proviennent de l'examen utilisant un mammographe, l'appareil dédié au dépistage du cancer du sein, peut être est-il le meilleur moyen existant pour établir un premier diagnostic et, ainsi confirmé la présence d'une pathologie mammaire, que ce soit des calcifications ou des masses, ce qui est plus dur encore à savoir, c'est si ces dernières sont cancéreuses donc « malignes » ou non cancéreuses « bénignes ». C'est ce que nous allons tenter, avec l'application que nous développerons, son but étant de savoir s' il y a pathologie et, si oui, si elle est maligne ou bénigne.

Tout au long de ce mémoire, nous allons évoquer les différents aspects que nous avons traité pour mener au mieux notre projet, que ce soit dans le contexte médical avec toute sorte de définition pour mieux comprendre le cancer du sein et, ainsi avoir une idée précise de ce que nous cherchons à établir, ou relativement à la segmentation d'images avec une fois encore, les définitions requises des algorithmes utilisés dans le cadre de notre application. La partie caractérisation, appréhendée suivant les différents paramètres que, l'on peut extraire d'une image pour l'analyser et, finalement la partie classification, nous aide à comprendre le processus de décision pour développer notre application présenté.



Le travail appréhendé dans ce projet, suscite l'intérêt du laboratoire LATSI (laboratoire du traitement du signal et de l'image de l'université Sâad Dahlab de Blida), depuis une décennie, pour la mise en œuvre d'un système de détection de pathologies mammaires; à savoir les calcifications, les masses et les cas sains. Notre problème central, a été la segmentation d'image s'adaptant aux types de structures recherchées ainsi qu'à leur caractérisation topologique et à leur classification, suivant une analyse connexionniste.

Notre objectif est alors, d'extraire les différentes structures composant l'aspect pathologique du sein, suivant une analyse texturale et morphologique et, une classification des différentes pathologies rencontrées selon, le réseau de neurones en coopération avec les algorithmes génétiques, pour l'aide au diagnostic clinique.

Ce mémoire est rédigé suivant l'organisation ci-dessous ;

- Le premier chapitre introduit le contexte mammographique ainsi, que quelques notions sur le cancer du sein,
- Le second chapitre est consacré aux principales techniques de la segmentation d'images, comme les contours actifs, l'analyse d'histogramme et la morphologie mathématique.
- Le troisième chapitre appréhende l'analyse morphologique et texturales et, spécifie les paramètres choisis pour la caractérisation des lésions mammographiques ainsi que la classification connexionniste et son association avec les algorithmes génétiques pour la prise de décision,
- Le quatrième chapitre est consacré à la présentation générale de la méthodologie adoptée ainsi qu'à, son application sur des images mammographiques d'une base de données réelles et, l'évaluation des performances du système développée.



Chapitre 1 Contexte mammographique et cancer du sein

Le cancer du sein, est le cancer le plus diagnostiqué chez les femmes à travers le monde, autant avant, qu'après la ménopause. Une femme sur 9, sera atteinte d'un cancer du sein au cours de sa vie et, 1 femme sur 27 en mourra.

Le plus souvent, le cancer du sein survient après 50 ans. Le taux de survie 5 ans après le diagnostic varie de 80 % à 90 %, selon l'âge et le type de cancer [1].

Le nombre de personnes atteintes a progressé régulièrement, au cours de trois dernières décennies. Par contre, dans les pays développés, le taux de mortalité a continuellement diminué au cours de la même période, grâce aux progrès réalisés en matière de dépistage, de diagnostic et de traitement. Cependant, dans les pays en voie de développement, le taux de survie est bas, car le dépistage précoce du cancer n'est pas encore suivi. Mentionnons que les hommes peuvent aussi en être touchés; ils représentent 1 % de l'ensemble des cas [2].

Dans ce chapitre, nous présentons l'approche médicale de l'imagerie du sein, pour situer le contexte de notre application. Nous décrivons ainsi, l'anatomie du sein, le cancer du sein et les changements bénins et malins qui peuvent se produire, sans pour autant être exhaustif, dans le but de faciliter la compréhension de la problématique, pour en déduire les méthodes adéquates à la détection des structures pathologiques.

1.1 Anatomie du sein

La fonction biologique du sein est de produire le lait maternel, afin de nourrir un nouveau-né. Le sein (appelé aussi glande mammaire) se compose de quinze à vingt lobes séparés par du tissu graisseux. Ces lobes sont, eux-mêmes, constitués de lobules capables, de sécréter du lait en période d'allaitement. Un lait est transporté par les canaux galactophores jusqu'au mamelon, situé au centre d'une zone pigmentée transcrite par l'aréole (cf. Figure 1.1).

La glande mammaire se développe et, fonctionne sous l'influence d'hormones produites en quantité variable, tout au long de la vie: les œstrogènes et la progestérone [3].

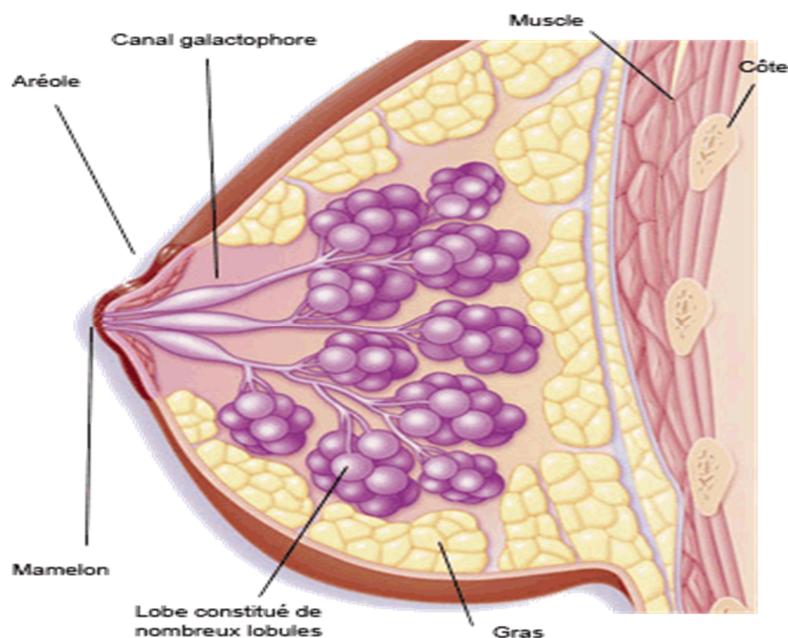


Figure 1.1 : structure du sein

1.2 Introduction au cancer du sein

Le cancer du sein (cf. Figure 1.2) est le cancer le plus fréquent chez la femme. Il s'agit d'un type de cancer qui se produit, lorsqu'il ya une croissance incontrôlée, de cellules anormales dans le sein. Le cancer du sein survient lorsque les cellules vieilles ou endommagées, se divisent et se multiplient sans contrôle. Cela se traduit, dans le développement d'une masse maligne du tissu (tumeur) dans le sein. Si elle n'est pas traitée, les cellules du cancer du sein peuvent continuer à se multiplier et s'étendre, à d'autres parties du corps lors, d'un processus appelé métastase [2].

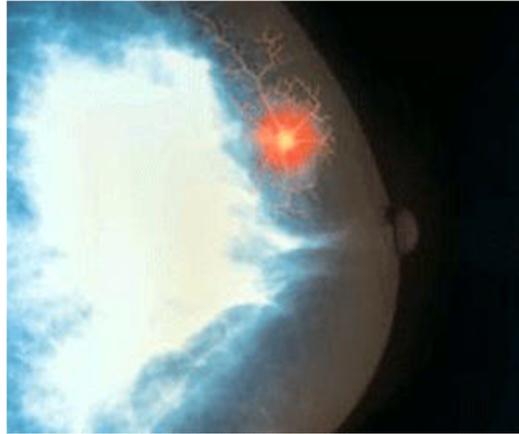


Figure 1.2 : sein cancérisé

Le pronostic du cancer du sein, varie selon le type et le stade d’avancement. Si le Cancer du sein est diagnostiqué et traité à un stade précoce, il est maîtrisable [7]. On distingue deux grandes catégories de tumeurs: les tumeurs bénignes et les tumeurs malignes (cf. Tableau 1).

Tumeur bénigne	Tumeur maligne
Bien limitée	Mal limitée
Histologiquement semblable au tissu d'origine	différenciation aberrante relativement au tissu d'origine
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 total
Pas de métastase	Métastase(s)
Encapsulée	Non encapsulée

Tableau 1.1 : critères de distinction entre tumeurs bénignes et malignes



1.2.1 Symptômes du cancer de sein

Un des symptômes caractéristiques d'une lésion mammaire, est l'apparition de grosseurs dans le sein. Dans la plupart des cas, ces grosseurs sont bénignes et ne constituent pas une tumeur, mais un kyste ou un adénome fibreux. Cependant, aucun symptôme n'est détecté au stade précoce du cancer. Une grosseur au sein qui est maligne, n'est généralement pas douloureuse, mais doit être rapidement, évaluée médicalement en raison du risque métastatique [1].

Lorsqu'une tumeur apparaît dans les canaux galactophores, du sang peut s'écouler du mamelon. La taille ou la forme du sein peut changer. En outre, le mamelon peut entrer vers l'intérieur ou la peau peut se rétracter, entraînant ainsi la formation d'une fossette. L'apparition d'une bosse ou d'une enflure dans l'aisselle, ou encore d'une rougeur ou d'un gonflement du sein sont d'autres symptômes. Ces transformations peuvent constituer des signes précoces du cancer du sein, mais elles peuvent aussi indiquer un autre trouble non cancéreux. En fait, seul le médecin pourrait repérer, précisément la cause exacte du nodule [2].

1.2.2 Facteurs de risque

Il n'y a pas une cause unique, ni un seul déclencheur, du cancer du sein. En fait, certains facteurs de risque accroissent la probabilité de cette maladie.

➤ *Facteurs internes*

- **l'âge** : trois cancers du sein sur quatre, se déclarent vers 50 ans. La maladie est rare chez la femme de moins de 35 ans et exceptionnelle, avant 20 ans. Pour cette raison, le programme de dépistage organisé du cancer du sein est bénéficiaire, aux femmes de plus de 50 ans, d'autant que la densité de la glande mammaire est moins importante à cet âge.
- **les antécédents familiaux** : 20 à 30 % des cancers du sein se manifestent chez des femmes présentant des antécédents familiaux (mère, sœur, etc.).
- **la prédisposition génétique** : 5 à 7 % des cancers du sein seraient liés à une prédisposition génétique héréditaire, c'est-à-dire existant déjà chez l'embryon.



- **les antécédents personnels** : les femmes ayant déjà eu un cancer du sein, présentent un risque 4 à 5 fois plus important d'en déclarer, un second [3].

➤ **Facteurs externes**

Les facteurs externes liés au cancer du sein, sont d'impact modéré. Le plus important est l'exposition de l'organisme aux hormones oestrogéniques, notamment au cours du traitement hormonal substitutif de la ménopause, le risque augmente, avec la durée. Nous citons aussi ; le tabac, l'alcool et le surpoids qui sont également considérés, comme des facteurs de risque, pour le cancer du sein. L'obésité multiplierait, ainsi par 1,5 à 2 fois la probabilité de déclarer la maladie.

Une femme sujette à l'un des facteurs de risque précités, peut ne jamais développer un cancer du sein. *A contrario*, il est possible qu'une femme n'ayant aucun de ces facteurs de risque, soit atteinte d'un cancer du sein [3].

1.2.3 Catégories du cancer de sein

Les divers types de cancers du sein (cf. figure 1. 3), évoluent de façon différente.

a. Cancer non invasif

Il s'agit d'un carcinome canalaire *in situ*. C'est le type le plus fréquent du cancer du sein, non invasif chez la femme. Comme son nom l'indique, il se forme à l'intérieur des canaux de lactation du sein. On diagnostique beaucoup plus fréquemment ce type de cancer, depuis l'utilisation plus répandue de la mammographie. Le traitement de ce cancer mène à la guérison dans presque tous les cas. Normalement, il ne se dissémine pas. Dans des cas exceptionnels, sans traitement, il poursuit sa croissance et peut alors devenir « infiltrant » donc se propager à l'extérieur des canaux de lactation [5].

b. Cancer invasif ou infiltrant

Ces formes de cancer envahissent les tissus autour des canaux de lactation, mais demeurent à l'intérieur du sein. Par contre, si la tumeur n'est pas traitée, elle peut se disséminer à

d'autres parties du corps (par exemple, les os, les poumons ou le foie) en générant des métastases.

1. Carcinome canalaire

Il se forme dans les canaux de lactation. Les cellules cancéreuses, traversent la paroi des canaux galactophores.

2. Carcinome lobulaire

Les cellules cancéreuses apparaissent dans les lobules regroupés dans les lobes. Puis, elles traversent la paroi des lobules et se disséminent dans les tissus environnants.

3. Carcinome inflammatoire

Un cancer rare qui se caractérise principalement par un sein qui peut devenir rouge, enflé et chaud. La peau du sein peut aussi prendre l'aspect d'une peau d'orange. Ce type de cancer progresse plus rapidement et est plus difficile à traiter.

4. Autres carcinomes (médullaires, colloïdes ou mucine, tubulaires, papillaires)

Ces types de cancer du sein, sont plus rares. Les principales différences entre ces types de cancer reposent sur le type de cellules touchées.

5. Maladie de Paget

Un cancer rare qui se manifeste par une petite plaie au mamelon qui ne guérit pas [5].

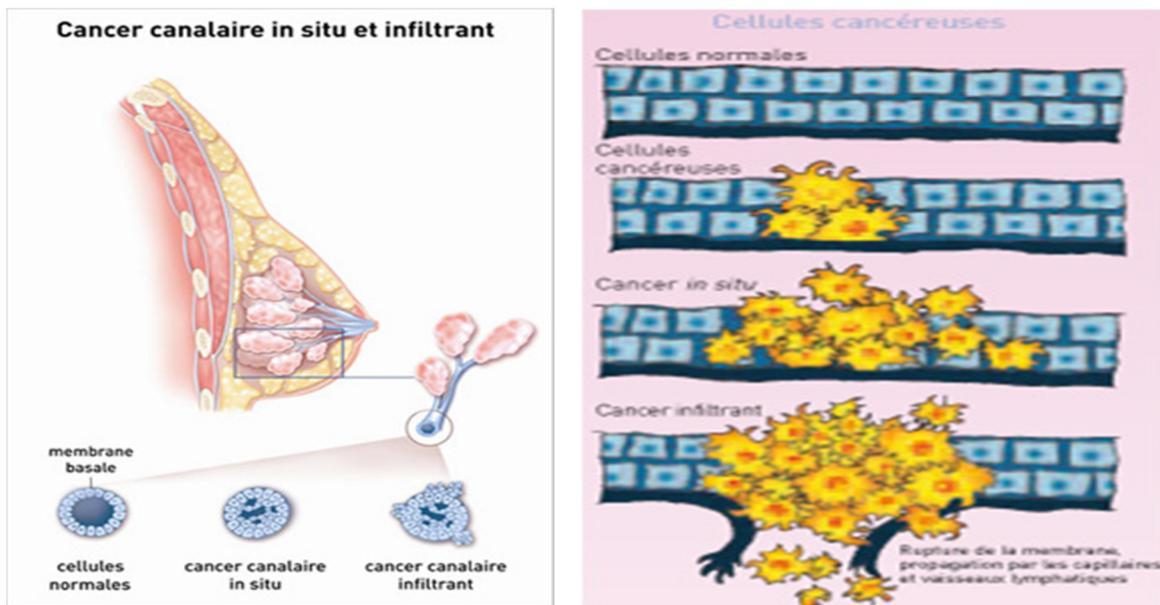


Figure 1.3 : différents types de cancer

1.3 Imagerie mammographique

L'imagerie médicale, est certainement l'un des domaines de la médecine qui a le plus progressé ces vingt dernières années. Ces découvertes, permettent non seulement un meilleur diagnostic mais offrent aussi, de nouveaux espoirs de traitement pour de nombreuses maladies, dont le cancer du sein. L'identification précise de la lésion, facilite le recours à la chirurgie, seule solution thérapeutique pour certains types. De telles techniques permettent alors, une aide à la décision.

1.3.1 Principe de la mammographie

Il s'agit d'un examen radiographique, qui permet de montrer la structure des seins afin de détecter d'éventuels masses (modifications des tissus du sein incluant, une perte de l'élasticité de certains éléments, appelée fibrose), kystes (Poche de liquide dont la peau est plus ou moins tendue) ou nodules. Ces masse, kystes et nodules, peuvent êtres à l'origine de "boules" sur les seins et/ou de douleurs. La mammographie est le meilleur moyen actuel, de dépistage du cancer du sein.

L'appareil de mammographie appelé, un mammographe (cf. Figure 1.4); se compose d'un tube à rayons X et d'un système de compression du sein [6].



Figure 1.4 : mammographe numérique

On distingue deux méthodes de reproduction de l'image mammographique à partir du film radiologique

- **numérisation indirecte** : s'effectue à partir d'un film et une camera CCD ou une numérisation du film à balayage laser (scanner optique). Cependant, une radiographie scannée, ne contiendra jamais plus d'information que l'épreuve originale. Elle conserve, tous les défauts de l'image de départ (bruit, dynamique, artefacts de développement [4].
- **Mammographie numérique** : Le couple écran film, est remplacé par un détecteur qui convertit proportionnellement, une intensité de rayons X, transmise par le sein, en un signal électronique. La gamme dynamique d'un détecteur numérique, est beaucoup plus grande que celle du couple écran film, ce que justifie la détection de certains détails de l'ordre du micron [4].

En mammographie, comme on peut le voir sur la figure 1.5, le sein est comprimé entre le support patient, sous lequel se trouvent le détecteur, et la pelote de compression. La compression est notamment appliquée pour limiter la dose délivrée à la patiente, pour réduire la quantité de rayonnement diffusé, pour étaler les tissus et pour immobiliser le sein. Un autre élément important du système de mammographie réside dans sa source de rayonnement, ou tube à rayons X, dont le foyer est positionné à la verticale du bord du détecteur le plus proche de la patiente, afin d'assurer une couverture complète du sein par le faisceau de rayons X qui en est issu.

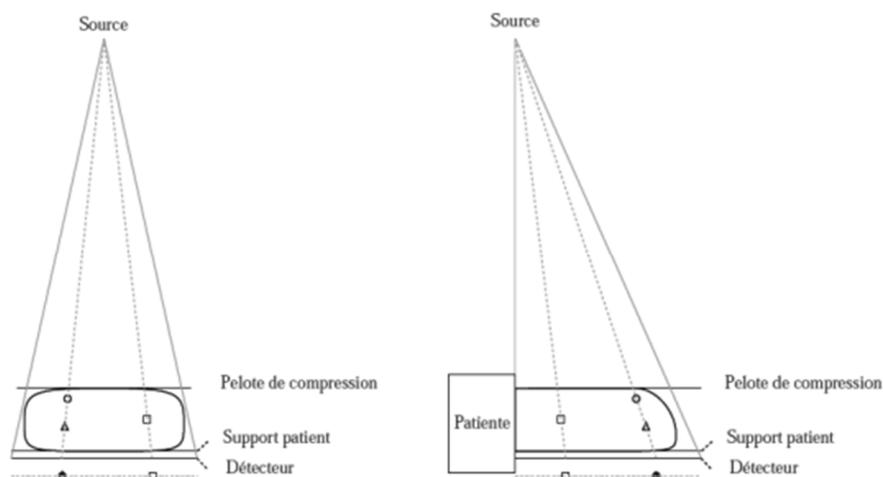


Figure 1.5: Géométrie d'acquisition d'une mammographie (vue de face et vue de profil).

L'examen typique de dépistage en mammographie comprend généralement l'acquisition d'images 2D sous deux incidences (cf. figure 1.6) ; incidence oblique externe et incidence crânio-caudale. Ces mammographies 2D sont réalisées selon différents angles d'incidence, pour minimiser les défauts d'identification de structures dues à la superposition de tissus qui pourrait, masquer des lésions ou au contraire en faire percevoir de fictives.



Figure 1.6 : clichés associés

1.3.2 Types de l'examen

a. Examen de dépistage

La mammographie, permet de déceler des tumeurs de très petite taille (les microcalcifications), bien avant qu'elles ne soient palpables ou qu'elles ne se manifestent par d'autres symptômes. Elle peut s'effectuer, dans le cadre d'un programme de dépistage organisé avec contrôle de qualité obligatoire, ou de manière individuelle [6].

b. Examen diagnostique

La mammographie de diagnostic, est un examen réalisé en cas de symptômes, au niveau des seins, détectés soit par la femme, soit lors d'une consultation médicale. La mammographie est dirigée, en particulier sur la zone suspecte et accompagnée d'autres examens (palpation, échographie, etc.) pour affiner l'analyse médicale. L'étape de diagnostic étant décisive (le traitement à prescrire est fortement lié à ce diagnostic), la lecture de la



mammographie est généralement assurée par deux radiologues. Dans le cas d'une lésion bénigne, un suivi à moyen terme, est recommandé. Ce suivi permet de vérifier si la lésion est stagnée ou évolutive. Par contre, dans le cas d'une lésion maligne, une ablation partielle ou totale, une chimiothérapie, une radiothérapie ou une hormonothérapie est préconisée voire l'association de plusieurs thérapies, est envisageable [7].

1.4 Pathologies mammaires

Après avoir décrit la démarche d'analyse d'un bilan mammographique, les différents types d'anomalies seront détaillées, en soulignant leur éventuelles particularités sémiologiques, ainsi que les explorations complémentaires utiles à leur caractérisation [6] [8].

Les anomalies détectées, se déclinent ci-dessous :

1.4.1 Kyste

C'est la "lésion" la plus fréquemment rencontrée. La taille d'un kyste peut varier de quelques millimètres à centimètres. Ils sont multiples et on les trouve des deux côtés. Plus de 30% des femmes dans la trentaine ont de multiples kystes de moins de 5mm. Cette fréquence est si importante que l'on peut s'interroger sur la légitimité du terme lésion. C'est pour cette raison que ce terme a été entouré de guillemets. Ces kystes sont des cavités remplies de liquide. Un kyste simple est virtuellement toujours bénin. Ces kystes se manifestent le plus souvent comme des masses mobiles.

1.4.2 Masses

Une masse (cf. Figure 1.7) est une lésion importante occupant un espace et, vue sur deux incidences différentes. Si une opacité potentielle est vue seulement sur une seule incidence, alors elle est appelée asymétrie jusqu'à ce que son caractère tridimensionnel soit confirmé.

Une masse, a tendance à être plus dense dans le milieu que vers les bords. Une masse n'est pas intercalée avec les cellules adipeuses à la même mesure, que les tissus non touchés. Une masse trouvée lors d'une mammographie, sera généralement être décrite selon sa forme, la densité et sa marge [7]. On distingue plusieurs de masses :

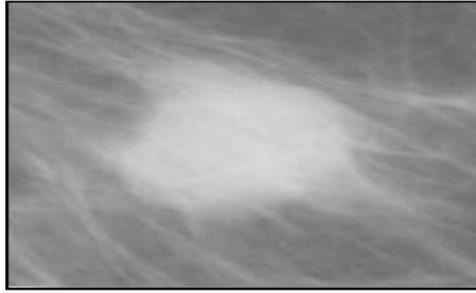


Figure 1.7 : Exemple d'une masse en mammographie

a. Distorsions architecturales

Les Distorsions Architecturales (DA) (cf. Figure 1.8) sont définies comme étant, un changement au niveau de l'architecture normale du sein, généralement observées dans les mammographies, comme une accumulation anormale de lignes droites qui, convergent vers un point central unique. Mais sans densité centrale radio-opaque. Souvent, ces DA peuvent être présentes avec les microcalcifications. Il est estimé que dans les dépistages de mammographie, entre 12 et 45% des cancers ne sont pas détectés à cause des DA [7].

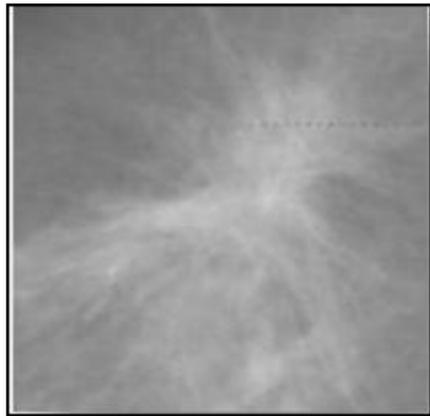


Figure 1.8 : exemple d'une distorsion architecturale

b. Asymétrie de densité

Les asymétries de densité (cf. Figure 1.9) sont le plus souvent bénignes. Elles correspondent soit à une répartition du tissu mammaire normal, soit à une symétrie focale. Elles peuvent être détectées, sur une ou deux incidences mammographique. En cas d'asymétrie focale, il ne faudra compléter le bilan standard par des cliché localisés, voire des agrandissements afin d'éliminer la présence de critère péjoratif [7].



Figure 1.9: exemple d'asymétrie : une région suspecte dans le sein droit, déduite d'une comparaison avec sein gauche

c. Lésions spéculées

Ces pathologies présentent des formes irrégulières (cf. Figure 1.10), des microlobulations nombreuses et petites (1-2 mm). Leur prévalence est faible, estimée à 0,30 % (11 cancers chez 3 627 patientes asymptomatiques avec seins denses) [5]. Une masse aux bords étoilés est très suspecte d'un cancer: les spicules représentent la modification tissulaire en rapport avec l'infiltration tumorale.

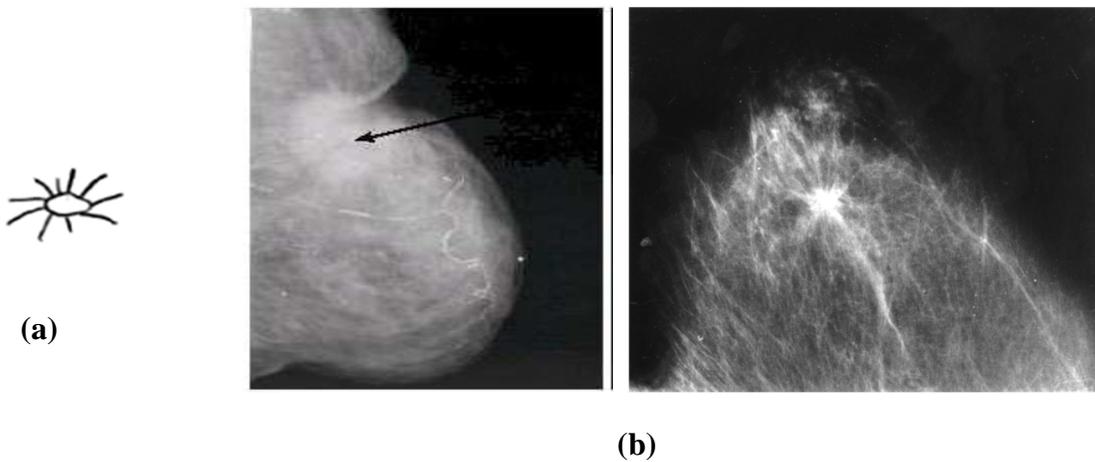


Figure 1.10: exemple de masses spéculées
(a) : forme d'une lésion spéculée
(b) : Opacités à contours spéculés avec rétraction cutanée traduisant un cancer à un stade avancé

d. Classification des masses mammaire [9]

Toute anomalie mammographique, doit être analysée de façon systématique. Les signes qui suivent, sont en faveur de la bénignité mais ne sont en aucun cas formels.

- **La taille** : elle est variable de quelque millimètre à plusieurs centimètres. Cependant, la taille ne prédit pas le caractère malin, sauf sur des clichés successifs lorsque l'on voit la taille augmentait régulièrement ;
- **La densité** : ces masses sont en général denses que le tissu mammaire adjacent ;
- **La forme** : on peut envisager cinq formes de masses (cf. Figure 1.11) :



Figure 1.11 : formes schématiques des masses

- **Les contours** : c'est critère pertinent pour l'étude des masses. La figure 1. 12, montre cinq types de contours:



Figure 1.12 : contours schématiques des masses

1.4.3 Calcifications

Ce sont des dépôts de sels de calcium qui apparaissent sous forme de points "blancs et brillants" à la mammographie; en fonction de leur taille, on parle de micro ou macro calcifications. Ces dernières, sont souvent bénignes alors que les microcalcifications demandent plus d'attention. Le diagnostic dépend de leur position dans le sein, de leur arrangement géométrique et du nombre de microcalcifications dans un amas et, aussi de la

comparaison avec une mammographie antérieure. On distingue 5 types de microcalcifications (cf. Figure 1.13), les calcifications de type I (rondes et planes), celles de type II (images régulières), celles de type III (images en grains de sel), les calcifications de type IV (points irréguliers) et celles de type V (formes vermiculaires) [8].

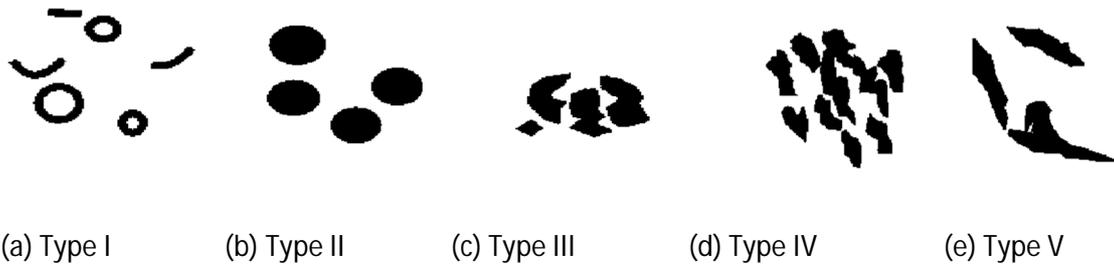


Figure 1. 13: différentes formes des microcalcifications

Les microcalcifications (cf. Figure 1.14) sont des petits dépôts de calcium de taille inférieure à 1 mm, qui Apparaissent comme des petits points lumineux situés dans la glande mammaire. Ces microcalcifications sont caractérisées, par leur type et les propriétés de leur distribution [5].

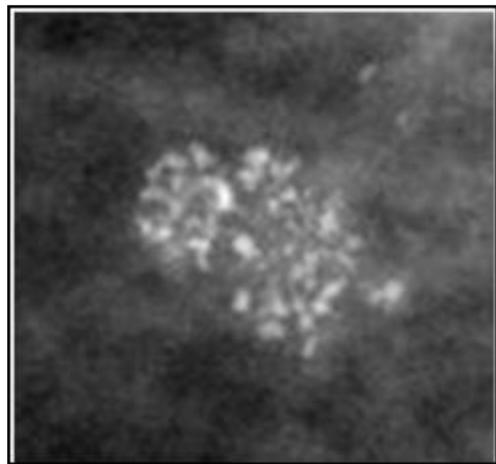


Figure 1.14 : exemple de microcalcifications

1.5 Caractéristiques radiologiques

1.5.1 Spécificité et sensibilité

La sensibilité est le pourcentage d'individus atteints d'une maladie donnée, qui auront un résultat positif lors d'un test de dépistage de la maladie, c'est-à-dire, ceux qui sont



identifiés comme étant, atteint de la maladie. Formellement, la sensibilité est la probabilité conditionnelle (positif/malade), elle représente le taux de faux négatifs.

La spécificité est une mesure de la capacité d'un test à donner un résultat négatif, lorsque la maladie est absente, C'est l'habileté du dépistage à bien identifier les femmes qui ne sont pas atteintes de cancer du sein. En ce sens, cette mesure nous permet de déterminer la proportion des femmes dépistées sans cancer, au stade préclinique qui sont correctement identifiées négatives par le dépistage.

La spécificité correspond à la probabilité que le dépistage produise un résultat négatif chez les femmes indemnes, du cancer. Par opposition, la proportion des femmes non porteuses du cancer chez qui la mammographie est positive, sont des résultats faussement positifs [8].

1.5.2 Faux positifs et faux négatifs

Faux positif : sur 5 à 7% des mammographies de dépistage, les radiologues concluent à une anomalie (examen positif). Si les investigations pratiquées à la suite du dépistage montrent que cette anomalie ne correspond pas à un cancer, ce qui arrive 9 fois sur 10, le résultat de la mammographie de dépistage est faussement positif. Ce phénomène est inévitable, car les signes d'alerte sont peu spécifiques, mais par leur formation et leur expérience les radiologues s'efforcent de le maintenir au taux le plus bas possible [8].

Faux négatif : résultat négatif de la mammographie de dépistage, qui s'avère par la suite erroné. La patiente ou son médecin, découvre un cancer du sein palpable, alors que la mammographie récente (moins de deux ans) n'a rien montré d'anormal (cancer d'intervalle). En réexaminant la mammographie de dépistage, on peut arriver à l'une des deux conclusions suivantes : il n'y avait effectivement pas de signe alarmant sur la mammographie (cancer occulte) ou un signe pouvant évoquer, un cancer a été manqué [8].



1.5.3 Densité mammaire

La densité mammaire [9], est un critère au moins aussi important, que l'âge pour la détermination du risque de cancer du sein. Ce résultat est issu de la plus grande étude sur le sujet, en termes de populations et du nombre de risques étudiés.

La densité mammaire est la mesure de combien de tissus, on peut voir sur la mammographie. Certains tissus, comme la glande mammaire, sont denses et apparaissent blancs sur l'image aux rayons X. La densité fait qu'il est difficile, pour les médecins de voir les tumeurs qui sont aussi de couleur blanche sur la mammographie. Les tissus gras, sont moins denses et sont de couleur claire: cela permet une meilleure détection des tumeurs.

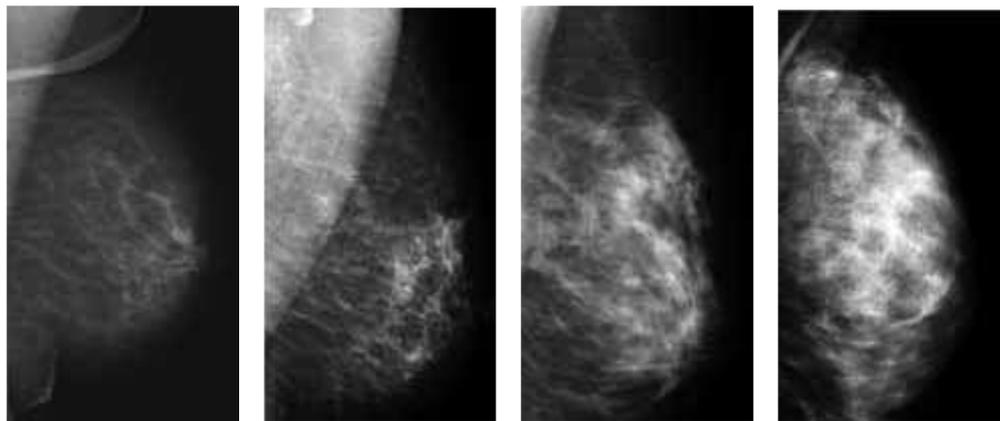
La densité mammaire est responsable de deux effets:

- une perte de sensibilité dans la détection de cancer (avec augmentation de cancer a l'intervalle) ;
- une perte de spécificité (avec une augmentation de nombre positif).

1.5.4 Classification de la densité mammaire [9]

Nous remarquons dans ce contexte, quatre types (cf. Figure 1.15):

- **type 1** : seins clairs quasi totalement graisseux ;
- **type 2** : il persiste quelques reliquats fibre-glandulaires ;
- **type 3** : seins denses de répartition hétérogène il persiste de la graisse, localement le sien est dense ;
- **type 4** : seins extrêmement denses de façon homogène, l'analyse mammographie est gênée Par la densité (absence complète de graisse rétro-glandulaire).



Type 1

Type 2

Type 3

Type 4

Figure 1. 15 : différents types de la densité mammaire

1.6 Conclusion

Nous avons vu dans ce chapitre, l'aspect clinique des principales pathologies mammaires, en passant par la technique de base de l'examen mammographique.

Les glandes mammaires peuvent être le siège, de tumeurs bénignes ou malignes. Le dépistage du cancer du sein, est vivement conseillé aux femmes dans le but d'une détection précoce de la pathologie, ce qui permettrait un traitement efficace. À partir de quarante ans, il est recommandé de faire régulièrement une mammographie, éventuellement complétée par une échographie.

La détection d'une anomalie en mammographie, repose tout d'abord sur le contrôle de qualité de clichés fournis. Cette étape systématique est primordiale, pour valider le contraste, le bon positionnement et, la compression optimale de la glande mammaire. La lecture comparative des clichés des deux seins, la comparaison avec les bilans antérieurs et, la connaissance de pathologies mammaires, doivent permettre au radiologue, de détecter des signes anormaux. Le plus souvent, des explorations complémentaires affinent l'examen. L'expérience montre, qu'il est plus facile de détecter une anomalie, que d'affirmer la normalité d'une mammographie.

Le chapitre suivant, est consacré à un parcours des techniques de segmentation en imagerie, afin d'opter pour une méthodologie de détection de pathologies mammographiques.



Chapitre 2 Techniques de segmentation en imagerie et méthodologie adoptée

2.1 Introduction

En analyse d'images, on distingue les traitements de bas niveau et les traitements de haut niveau. Cette distinction est liée au contenu sémantique des entités traitées et extraites de l'image. Les traitements de bas niveau opèrent, en général, sur les grandeurs calculées à partir des valeurs attachées à chaque point de l'image. Les traitements de bas niveau opèrent plutôt sur des données de nature numérique. A l'opposé, les traitements de haut niveau, s'appliquent à des entités de nature symbolique associées à une représentation de la réalité extraite de l'image. La segmentation d'images est l'un des problèmes phares du traitement d'images: elle consiste à partitionner l'image en un ensemble de régions connexes. L'intérêt de ces régions est de pouvoir être manipulé via des traitements de haut niveau, pour extraire des caractéristiques de forme, de position, de taille, etc....

La segmentation est l'opération fondamentale en analyse d'image [8], qui permet de détecter les contours ou les régions des objets. Dans ce qui suit, nous présentons quelques méthodes de segmentation, en choisissant des chemins de conception qui se basent soit, sur une analyse des propriétés spatiales où l'on utilise les différents algorithmes présentés, pour extraire une région ou un contour; en passant brièvement, par quelques transformations de la morphologie mathématique, pour la détection de certaines structures, dans l'image mammographique. Quelques approches adoptées dans l'imagerie mammographique, seront introduites.



2.2 Principe de la segmentation

La segmentation est une tâche très importante et difficile, car elle dépend de la nature de l'image et de l'objectif de l'analyse. La littérature [8], propose de nombreuses techniques de segmentation, mais la plupart d'entre elles, ont besoin de plusieurs paramètres dont le réglage, qui exige souvent l'expertise humaine.

Il n'y a pas de méthode unique de segmentation d'une image, le choix d'une technique, est lié à la nature de l'image et aux primitives à extraire.

La segmentation [8] est un traitement de bas niveau qui consiste à créer une partition d'une image **A** en sous-ensembles **R_i**, appelés régions, tels que:

$$\begin{aligned} \forall i R_i &\neq \emptyset \\ \forall i, j; i \neq j R_i \cap R_j &= \emptyset \quad \dots\dots\dots 2.1 \\ A &= \bigcup_i R_i \end{aligned}$$

Une région est un ensemble connexe de pixels ayant des propriétés communes (**intensité, texture, ...**) qui les différencie des pixels des régions voisines. Lorsque le nombre de régions est limité à deux, la segmentation prend le nom de **binarisation** et, se résume alors à séparer un objet du fond. Dans certains cas, le nombre de régions est connu d'avance ainsi que certaines caractéristiques de ces régions. Pour ces situations, la segmentation prend le nom de classification et consiste à associer chaque pixel, de l'image à une de ces régions.

Globalement la segmentation d'une image *I* consiste à partitionner cette image en plusieurs régions $\{R_i\}_{i=1,\dots,k}$ non vide tel que l'intersection entre deux régions différentes, soit vide et l'ensemble des régions recouvre toute l'image. Une région est un ensemble de pixels connexes ayant des propriétés communes qui les différencient, des pixels des régions voisines.



La segmentation est une étape importante pour l'extraction des informations qualitatives de l'image. Elle fournit une description de haut niveau, chaque région est connectée à ses voisines dans un graphe et chaque région porte une étiquette donnant des informations qualitatives comme sa taille, sa couleur, sa forme, son orientation. L'image se réduit donc à un graphe de nœuds étiquetés qui contient presque toute l'information utile au système. Fondamentalement, la segmentation est un processus qui consiste à découper une image en régions connexes présentant une homogénéité selon un certain critère, comme par exemple la couleur. L'union de ces régions doit redonner l'image initiale [8].

Dès lors, la segmentation sera valide si les conditions suivantes sont vérifiées :

- ❖ Tout les composants de l'image doivent être classés, autrement dit chaque pixel doit appartenir à une classe;
- ❖ Les régions doivent être homogènes;
- ❖ Les différentes régions doivent être séparées (disjointes).

A chacune de ces régions, doit correspondre un objet dans l'image car, dans ce procédé d'analyse d'images, l'objectif ultime est d'être en mesure de décomposer une image en un groupe d'objets distincts la composant. En général, ces objets ont des propriétés qui leur sont propres par rapport à l'image elle-même. Ainsi, il est possible de distinguer de tels objets par différentes mesures telles:

- leur aspect connexe;
- leur couleur cohérente;
- leurs contours;
- leur texture;
- des informations a priori.

Bien qu'il existe une multitude d'algorithmes de segmentation selon le domaine et les contraintes étudiées, quatre grandes catégories peuvent se distinguer: la segmentation par seuillage, par régions, par contours et enfin la segmentation par coopération région-contour [9].

L'objectif de la détection de contours, est la détermination des points de l'image séparant des zones de caractéristiques différentes. Chaque méthode de détection de contours, est



justifiée de manière plus ou moins explicite, par la définition du type de transitions recherchées [8].

De manière duale, le but de la segmentation en régions, est de partitionner l'image en ensembles de points connexes possédant des propriétés d'homogénéité.

2.3 Différentes méthodes de segmentation

Il n'y a pas une méthode unique de la segmentation d'une image, le choix d'une technique est lié au premier lieu à la nature d'une image (présence de bruit, contours,...), et en second lieu au domaine d'application.

Nous allons donc présenter dans cette section, diverses techniques connues de la segmentation. On regroupe de façon usuelle, les méthodes de segmentation en quatre groupes basés respectivement sur [8] :

- la segmentation basée sur les régions;
- la segmentation basée sur les contours;
- la segmentation basée sur une approche globale de l'image, par exemple le seuillage par histogramme ;
- la segmentation basée sur la coopération entre les trois premières segmentations.

2.3.1 Approche région

Les méthodes de l'approche région, cherchent à regrouper directement des pixels ayant une propriété commune. L'ensemble des regroupements de pixels définit à la fin, une segmentation de l'image en régions connexes [8] [9].

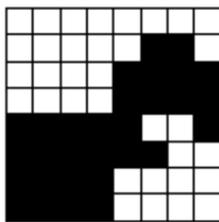
Il existe principalement quatre approches de la segmentation en régions :

- la segmentation par division de régions,
- la segmentation par fusion de régions,
- la segmentation par division-fusion de régions,
- la segmentation par croissance de régions.

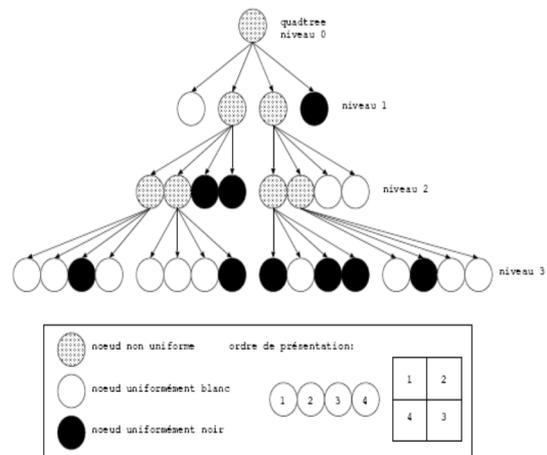
a. Segmentation par division de régions

L'approche de la segmentation par division de régions, consiste à diviser l'image originale en régions homogènes au sens d'un critère donné. Ce processus est récursif et, considère que la région initiale, correspond à l'image à analyser. Si une région ne respecte pas un prédicat d'homogénéité, elle est divisée en quatre sous régions de taille égale. Chaque sous région est ensuite analysée. L'algorithme récursif s'arrête, lorsque toutes les régions respectent le prédicat d'homogénéité.

Cette structure de données est un arbre quaternaire qui permet de stocker l'image à plusieurs niveaux de résolution. On part d'une région initiale qui est l'image tout entière. Si cette image vérifie un critère d'homogénéité, l'algorithme s'arrête. Sinon, on découpe cette région en quatre parties de même taille et on lance la procédure récursivement dans ces quatre parties. La région initiale va être stockée comme un nœud dans un graphe et les sous parties comme des fils de ce nœud. La figure 2.1.a montre, une image en noir et blanc 8x8 et le découpage correspondant à chaque niveau. La structure d'arbre associée à ce découpage est illustrée, dans la figure 2.1.b.



(a) : image 8x8



(b) : découpage de l'image 8x8

Figure 2.1 : arbre quaternaire sur une image 8x8 [8]

b. Segmentation par fusion de régions

C'est une structure de données constituée d'un graphe non orienté dont chaque nœud, représente une région et chaque arête représente une adjacence entre deux régions

[9]. La figure 2.3 représente, à titre d'exemple, le graphe d'adjacence de régions obtenu à partir de l'image de la figure 2.2.

L'analyse du graphe d'adjacence de régions permet de fusionner des régions d'une image sursegmentée. Le procédé consiste, à fusionner deux nœuds reliés par une arête à condition qu'ils respectent un critère de fusion.

Les méthodes d'analyse des graphes d'adjacence de régions, se distinguent selon l'ordre de parcours des différents nœuds du graphe et selon les critères de fusion. Comme pour la croissance de régions appliquée à l'image, l'ordre de balayage du graphe d'adjacence de régions influe sur les résultats de la fusion.

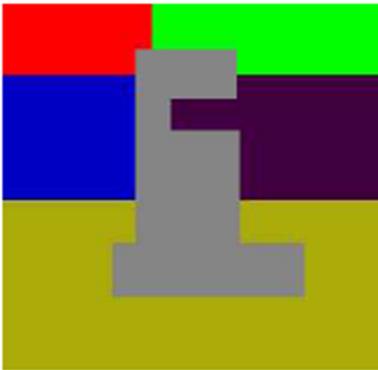


Figure 2.2 : image originale [8]

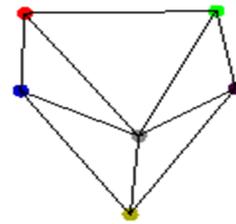


Figure 2.3 : graphe d'adjacence [8]

c. Segmentation par division fusion

La segmentation par division fusion, regroupe les deux types d'approches précédentes. Tout d'abord, l'image est divisée en régions homogènes qui respectent des critères globaux, puis les régions adjacentes qui, répondent à des critères locaux sont fusionnées [8].

Cette technique fonctionne en considérant, une représentation pyramidale de l'image analysée. Les régions sont de formes carrées et, correspondent aux différents éléments d'un étage donné de la pyramide suivante:

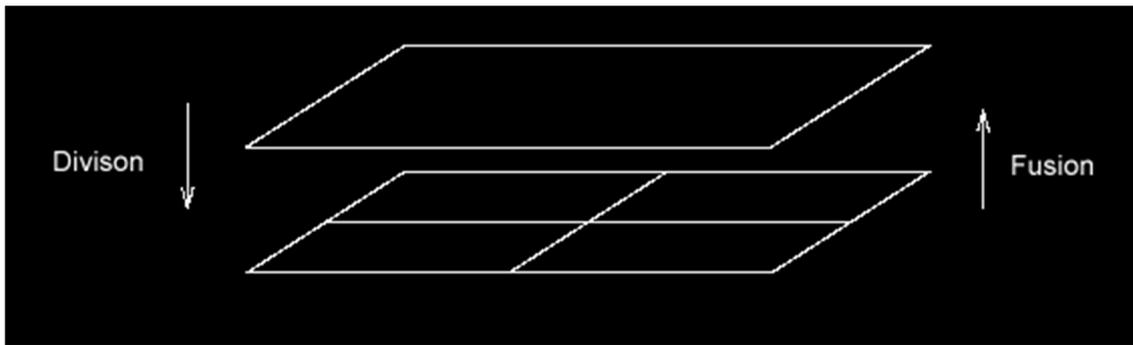


Figure 2.4: Illustration du concept pyramidal de l'image pour la segmentation par fusion et division de régions [9]

Ainsi, si une région de l'image considérée ne respecte pas les critères fixés de seuils, elle sera divisée en quatre nouvelles régions. Réciproquement, si 4 régions indépendantes mais voisines, respectent tous les critères de division, elles seront fusionnées dans l'étage supérieur de la pyramide.

En somme, cette segmentation peut se rapporter à la construction d'un 'quadtree' de segmentation où chaque feuille; représente une région dite homogène, c'est-à-dire qu'elle respecte les critères de seuil fixés.

d. Segmentation par croissance de régions

Ce type de segmentation, consiste à faire croître des régions, en y ajoutant successivement les pixels adjacents qui satisfont un critère d'homogénéité. La croissance s'arrête lorsque tous les pixels ont été traités [8] [9].

L'étape initiale, consiste à sélectionner les germes des régions qui correspondent généralement à un pixel. Puis, les régions sont construites, en y ajoutant successivement les pixels qui leur sont connexes et qui vérifient un critère de similarité.

La plupart des algorithmes de croissance de régions sont régis par un parcours de l'image selon le balayage de gauche à droite et de haut en bas.

L'avantage de la croissance de régions est de préserver la forme de chaque région de l'image [8] [9].

L'idée derrière cette méthode est de se fixer un point de départ dans l'image, que l'on nommera *germe* de la région. On fixe ensuite un critère d'homogénéité pour la région qui

est en général une intensité comprise entre deux valeurs, tout comme la méthode de seuillage. Par une procédure récursive, on inclut dans cette région, les points connexes qui vérifient le critère. De cette manière, on fait croître la région tant que le critère est respecté et on obtient en bout de ligne, une nouvelle région connexe.

Nous présentons ici une méthode largement utilisée sous la forme, d'un pseudo-code. Cette méthode récursive se nomme : l'algorithme du labyrinthe:

- on part d'un point aléatoire de l'image ;
- on marque ce point, comme faisant partie de la région en cours;
- on vérifie pour chaque voisin, s'il vérifie le critère d'homogénéité et, s'il n'a pas déjà été vérifié; si oui, on applique la même procédure au voisin, sinon, on arrête la procédure.

Cet algorithme possède plusieurs avantages dans le contexte de la segmentation:

- il assure de toucher tous les points d'un domaine connexe qui vérifient le critère ;
- c'est une méthode de 'force brute' puissante et simple.

Nous considérons à cet effet, l'application de cette méthode sur des images mammographiques (cf. Figure 2.5), selon les travaux de C.-H. Wei*, S.Y Chen, and X. Liu [10]:

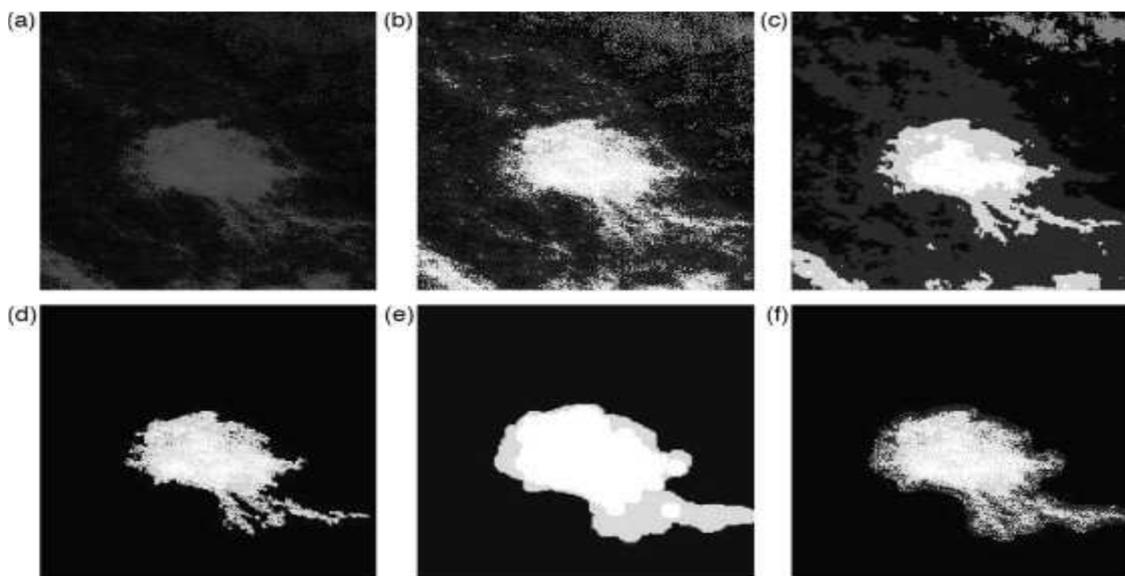


Figure 2.5 : application de l'algorithme croissance de région sur les images mammographiques [10]

2.3.2 Segmentation par les K-moyennes

Cette méthode [11] consiste à rassembler les pixels en K groupes (clusters). K étant un paramètre Préfixé qui détermine le nombre de régions. Et après avoir déterminé les paramètres (couleur Par exemple) de toute région, chaque point est affecté au centre le plus proche, pour recalculer Ensuite le paramètre de chaque région, jusqu'à ce qu'elles soient stables

Les principales étapes de cet algorithme sont :

- Choix aléatoire de la position initiale des K clusters.
- (Ré-) Affecter les objets à un cluster suivant un critère de minimisation des distances (Généralement selon une mesure de distance euclidienne).
- Une fois tous les objets placés, recalculer les K centroïdes.
- Répéter les étapes 2 et 3 jusqu'à ce que plus, aucune réaffectation ne soit faite.

La figure 2.6, illustre un exemple de l'application des K-means sur des images mammographiques, suivant les travaux de Patel et al [11].

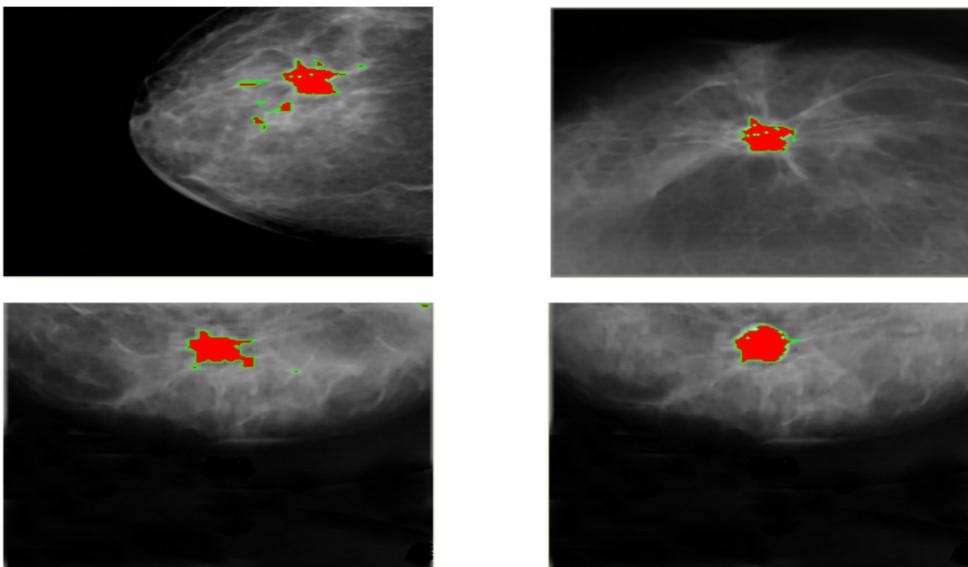


Figure 2.6: application de l'algorithme k-means sur les images mammographiques [11]

Discussion

Le principal avantage de cette méthode, est que le nombre de régions est connu au préalable (il n'y aura pas de problème de sur-segmentation ou sous-segmentation), mais l'inconvénient réside, en la difficulté de déterminer le nombre de clusters et, l'incohérence des régions (deux objets éloignés, peuvent appartenir à la même région).



2.3.3 Segmentation par analyse d'histogramme [12]

Il s'agit de méthodes, de base en segmentation d'image. Le principe général du seuillage, consiste à chercher une valeur appropriée du seuil, puis de classer tous les pixels de l'image selon la valeur de leurs niveaux de gris par rapport à ce seuil, afin de séparer les régions d'intérêt et l'arrière plan de l'image. Plusieurs méthodes de détermination de cette valeur de seuil, ont été appliquées pour la segmentation d'image mammographique. Certaines méthodes de détermination du seuil sont basées sur l'utilisation de paramètres autres que le niveau de gris, tels que l'entropie. A titre d'exemple, Otsu a formulé le problème du seuillage d'une image, comme étant un problème d'analyse discriminante itérative, qui permet de choisir une valeur optimale du seuil.

Le critère utilisé pour le choix du seuil est basé, sur la maximisation d'une mesure statistique de séparation entre les classes. Dans tous les cas, le seuil obtenu via les méthodes citées ci-dessus, sert finalement à la classification des pixels de l'image selon leurs niveaux de gris.

D'une façon générale, les méthodes de seuillage peuvent être classées en deux catégories :

a. Les méthodes de seuillage global

Ces méthodes sont largement utilisées en segmentation d'images mammographies, afin de détecter les zones tumorales. Le principe de ces méthodes, consiste à déterminer la valeur du seuil, en utilisant l'information globale contenue dans l'image. Cette information est souvent donnée sous forme d'histogramme des niveaux de gris contenus dans l'image. Malgré cette large utilisation, le seuillage global, s'avère peu efficace pour identifier avec précision les régions d'intérêt. En effet, l'image mammographique constitue, en réalité, la projection d'une scène 3D dans un espace d'observation 2D. Cette projection engendre des chevauchements importants des régions qui composent le tissu mammaire, ce qui limite l'efficacité de ces méthodes.

b. Les méthodes de seuillage local

Ces méthodes visent, à affiner localement la valeur du seuil pour mieux identifier les régions d'intérêts. La valeur du seuil, est déterminée en se limitant aux informations contenues, dans le voisinage local de chaque pixel. Ces méthodes, ont souvent montré une meilleure efficacité de détection, par rapport aux méthodes de seuillage global.



Nous présentons dans ce qui suit, d'autres méthodes de segmentation en régions ainsi que, quelques transformations de la morphologie mathématique, usées dans notre système.

2.3.4 Segmentation par contour

L'approche contour est l'approche duale de l'approche région: elle consiste à rechercher les frontières des régions caractérisées par une discontinuité visuelle. Ces méthodes sont basées sur la recherche de variations locales significatives. Largement étudiée pour les images en niveaux de gris, la recherche de contours repose principalement sur le calcul des dérivées [8] [13].

Pour une image monochrome, le passage d'une région à une autre, se traduit par un changement rapide de niveaux de gris (un gradient élevé dans la direction normale au contour). Le gradient se caractérise par sa direction et, sa norme.

Pour déterminer les positions de forts gradients, soit on calcule la dérivée première pour ne conserver que les régions de plus forte pente, soit on calcule les passages par zéro de la dérivée seconde. Les contours ainsi obtenus, forment rarement des frontières fermées, il faut par la suite appliquer, des algorithmes de fermeture de contours.

Les méthodes de segmentation basées sur l'approche contour, ont donc pour objectif de trouver les lieux de fortes variations du niveau de gris. Un nombre important de méthodes a été développé. Ces méthodes s'appuient sur la détection des discontinuités dans l'image et, peuvent être divisées en trois classes :

- les méthodes dérivatives;
- les méthodes analytiques;
- les méthodes basées sur les contours actifs.

On distingue trois types de contours simples (cf. Figure 2.7):

- marche d'escalier: le contour est net (contour idéal) ;
- rampe: le contour est plus flou;
- toit : il s'agit d'une ligne sur un fond uniforme.

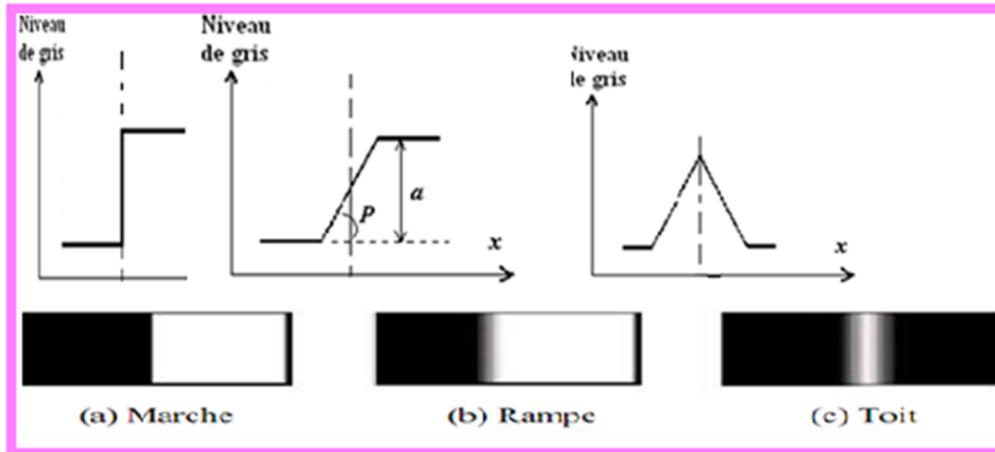


Figure 2.7 : Exemples de contours. a) Contour idéal. b) rampe. C) contour réel.

a. Segmentation par contours actifs [14]

Un contour actif est un ensemble de points qu'on va tenter de déplacer pour leur faire épouser une forme. Il s'agit d'une technique d'extraction de données utilisée en traitement d'images.

La courbe représentant le contour est défini sous la forme paramétrique de l'équation suivante :

$$C = (x(s,t), y(s,t)) \tag{2.2}$$

L'idée de cette méthode, est de déplacer les points pour les rapprocher des zones de fort gradient tout en conservant des caractéristiques comme la courbure du contour ou la répartition des points sur le contour ou d'autres contraintes liées à la disposition des points.

Au démarrage de l'algorithme, le contour est disposé uniformément autour de l'objet à détecter puis il va se rétracter pour en épouser au mieux ses formes. De la même manière, un contour actif peut aussi se dilater et tenter de remplir une forme, il sera alors situé à l'intérieur de celle-ci au démarrage de l'algorithme.

A chaque itération, l'algorithme va tenter de trouver un meilleur positionnement pour le contour pour minimiser ses dérives par rapport aux contraintes utilisées. L'algorithme



s'arrêtera, lorsqu'il ne sera plus possible d'améliorer, le positionnement ou simplement, quand le nombre maximum d'itérations aura été atteint. On utilise les notions d'énergies interne et externe pour caractériser respectivement; la forme du contour et tous les éléments qui lui sont propres et, le positionnement du contour sur l'image en tenant compte des lignes de gradient.

➤ Phases de l'algorithme

L'algorithme suit les phases ci-dessous à chaque itération:

- calcul des énergies interne et externe, caractérisant le contour lui-même et son positionnement sur l'image;
- pour chaque point du contour, détermination d'une nouvelle position, sur laquelle le contour devrait mieux minimiser les écarts de contraintes;
- arrangement du contour pour qu'il respecte des contraintes d'écartement entre les points, de régularité de points,...etc.

➤ Energie interne

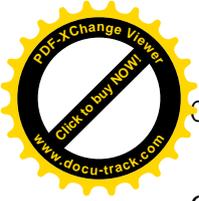
L'énergie interne ne dépend pas de l'image ni de la forme à détourner, elle ne dépend que des points du contour. Elle regroupe des notions comme la courbure du contour ou la régularité d'espacement des points. En effet, le contour doit conserver une forme arrondie en minimisant les dérivées d'ordre 1, 2, ... et doit empêcher un point de se détacher trop loin du reste du contour.

Idéalement, l'énergie interne est minimale pour un cercle où tous les points sont régulièrement espacés.

➤ Energie externe

L'énergie externe correspond à l'impact du contour sur l'image. Pour la calculer, il faut considérer l'opposé de la valeur de son gradient (ou de toute autre représentation mettant en jeu les contours à épouser) en chaque point du contour.

Cette énergie externe doit théoriquement être minimale si le contour épouse parfaitement la forme à extraire.



On ne considère l'opposé du gradient que, pour avoir une énergie externe minimale à la convergence de l'algorithme.

On pourra utiliser une représentation en gradient et en flot pour ajouter de l'information dans les zones uniformes où le gradient est nul, de manière à guider le snake vers le bord le plus proche.

➤ **Utilisation des deux énergies**

Chaque position du contour actif donne une énergie interne et une énergie externe dont la somme doit être minimisée et influencera les mouvements des points du contour actif. La grande difficulté de l'utilisation des snakes, réside dans le choix des pondérations à donner à chaque énergie.

Traditionnellement il est d'usage d'utiliser un paramètre pour l'énergie interne et un autre pour l'énergie externe, mais un réglage plus fin peut s'opérer en ajoutant des paramètres différents pour des aspects différents de l'énergie interne ou de l'énergie externe.

On pourra ainsi avoir :

- un paramètre pour pondérer l'énergie dégagée par la mauvaise courbure du snake;
- un autre pour pondérer l'énergie dégagée par le non respect de l'écartement régulier entre ses points;
- un autre pour pondérer l'énergie dégagée par le positionnement sur les lignes de gradient de l'image.

➤ **Modifications du contour**

Après avoir calculé l'énergie globale dégagée par le contour et par son positionnement sur l'image, il convient de déterminer comment le faire évoluer pour minimiser cette énergie. Pour cela, une méthode simple et intuitive est d'observer les pixels voisins immédiats de chaque point du contour pour déterminer pour chacun d'eux l'énergie globale du snake, chaque meilleur voisin devenant un point du contour.

Il est nécessaire que le contour possède toujours suffisamment de points pour être sûr de bien calculer son énergie globale, en particulier son énergie externe qui sera plus précise en tenant compte d'avantage de points. C'est la raison pour laquelle il peut s'avérer pertinent de rajouter ou de supprimer des points à chaque itération si des contraintes ne sont pas

suffisamment respectées. Par exemple, on pourra rajouter un point au snake, si ses voisins sont trop éloignés. A l'inverse, on pourra supprimer un point s'il est trop près de ses voisins. La figure 2.8, illustre un exemple de l'application des contours actifs.

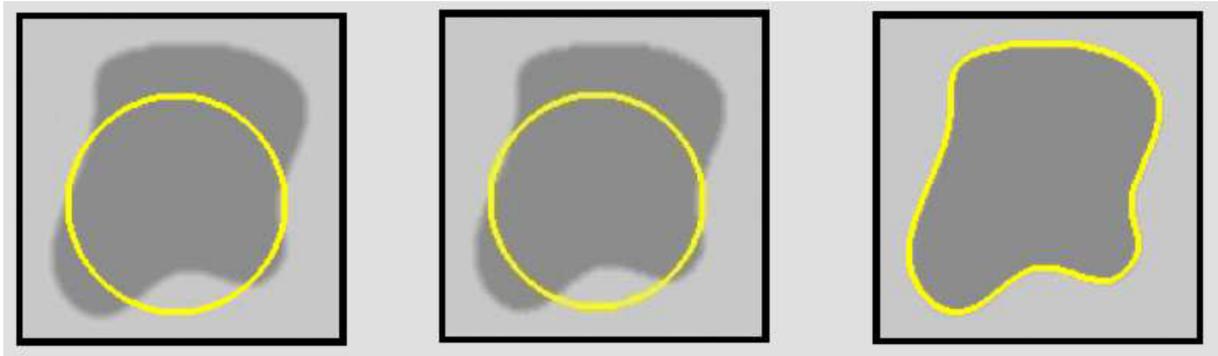


Figure 2.8 : illustrations des grandes étapes du déroulement d'un contour actif

b. Segmentation par les ensembles de niveaux (levels sets) [15]

La méthode des ensembles de niveaux, est une méthode de simulation numérique utilisée pour l'évolution des courbes et des surfaces dans les domaines discrets. C'est une représentation variationnelle des contours, qui conduisent à des solutions qui évoluent au cours du temps dans l'image, régies par un critère global.

Cette méthode permet de faire évoluer une courbe paramétrique fermée $C(p)$ suivant une équation du type :

$$\frac{\partial C}{\partial t} = FN \quad 2.3$$

Où t est le temps, F est la vitesse d'évolution et N est la normale unitaire à la courbe.

Chaque point de la courbe C est évalué, suivant la direction normale à la courbe avec une vitesse F .



L'avantage principal de cette méthode, est la possibilité de gérer automatiquement le changement de topologie de la courbe en évolution. La courbe C peut être divisé en deux ou trois courbes. Inversement, plusieurs courbes peuvent fusionner et devenir une seule courbe.

La méthode des Ensembles de Niveaux, est capable de réaliser ce changement en utilisant une fonction d'ordre supérieure :

$$\phi : R^{N+1} \rightarrow R^N \quad 2.4$$

C étant le niveau zéro de la courbe d'ordre supérieur. Il est représenté selon:

$$C(t) = \{s \in R^2 \mid \phi(s, t) = 0\} \quad 2.5$$

Le niveau zéro (C) peut se diviser, tandis que, la fonction ϕ reste continue.

➤ **Principe de la méthode**

Le principe des contours actifs est de faire évoluer une courbe. On a vu dans la section précédente une formulation dynamique qui fait intervenir la vitesse d'évolution du contour. On va donc s'intéresser à la façon de faire évoluer la courbe et plus généralement à la notion de propagation de fronts.

➤ **Avantages**

1. Prise en compte des changements de topologies automatiques grandeurs géométriques intrinsèques (normales entrante/sortante, courbure) faciles à calculer;
2. extension a la 3D simple: il suffit d'ajouter une coordonnées a l'équation d'évolution de la fonction Φ : on a alors un volume $\Phi(x, y, z, t)$;
3. Discrétisation de Φ avec une grille définit dans le domaine de l'image (valeurs possibles pour (x, y));
4. Utilisation des méthodes numériques connues pour calculer les dérivées.

➤ **Inconvénients**

L'implémentation implique plusieurs problèmes:



1. On doit construire une fonction initiale $\Phi(x, y, z, t=0)$ de manière à ce que, son niveau zéro corresponde, à la position initiale du contour;
2. L'équation d'évolution n'est dérivée qu'au *niveau* zéro, la fonction vitesse v n'est donc pas définie (en générale) pour les autres *niveaux*;
3. La déformation constante, peut causer la formation de coins saillants sur le modèle initialement lisse. Ce coin, peut perturber les déformations successives, puisque la définition de la normale devient ambiguë.
4. il est parfois nécessaire, de recalculer la fonction distance par rapport au niveau zéro.

2.3.5 Approche coopérative [16] [17]

Plusieurs chercheurs s'accordent pour affirmer que la coopération entre les segmentations par régions et par contours, contribuent à une meilleure prise en compte des caractéristiques des entités de l'image et, par conséquent, à une meilleure segmentation.

En effet, les algorithmes combinant les techniques de segmentation basées sur les régions et celles basées sur les contours prennent avantage de la nature complémentaire de l'information sur la région et sur le contour. Ainsi, une segmentation par coopération régions-contours peut être exprimée, comme une entraide entre ces deux concepts afin d'améliorer le résultat final. Elle exploite les avantages de ces deux types de segmentation, pour aboutir à un résultat de segmentation plus précis et plus fidèle que celui obtenu à l'aide d'une seule technique.

La segmentation d'images par coopération régions-contours peut se faire en fonction de trois approches:

a. Coopération séquentielle

Il s'agit d'une coopération (cf. Figure 2.9) dans laquelle l'une des techniques de segmentation (région ou contour) est réalisée en premier lieu; son résultat va être exploité par l'autre technique pour renforcer la définition des critères ou des paramètres de la segmentation.

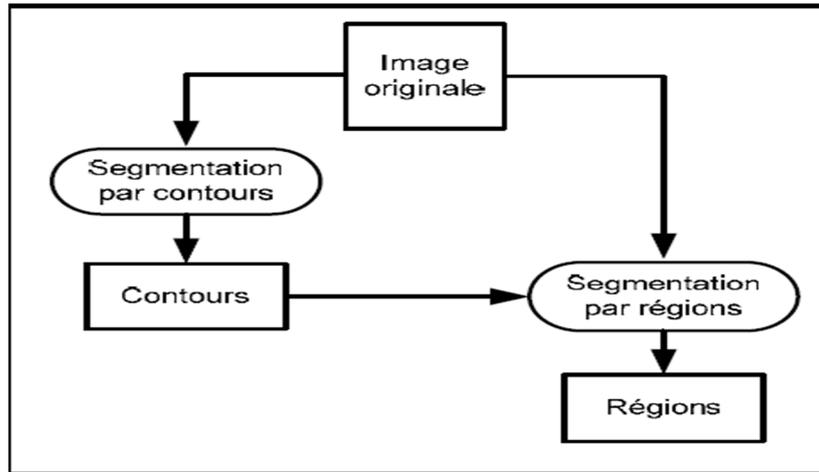


Figure 2.9 : schéma de la coopération séquentielle

b. Coopération des résultats

Les deux types de la segmentation seront réalisés indépendamment; la coopération (cf. Figure 2.10) concernera leurs résultats qui seront intégrés afin d'atteindre une meilleure segmentation;

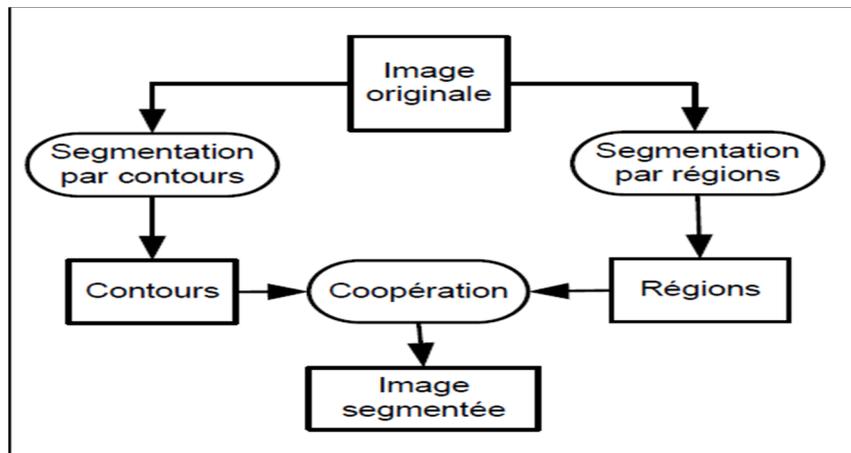


Figure 2.10 : synoptique de la coopération des résultats

c. Coopération mutuelle

Les deux types de segmentations coopéreront mutuellement (cf. Figure 2.11) au cours de leur processus d'exécution.

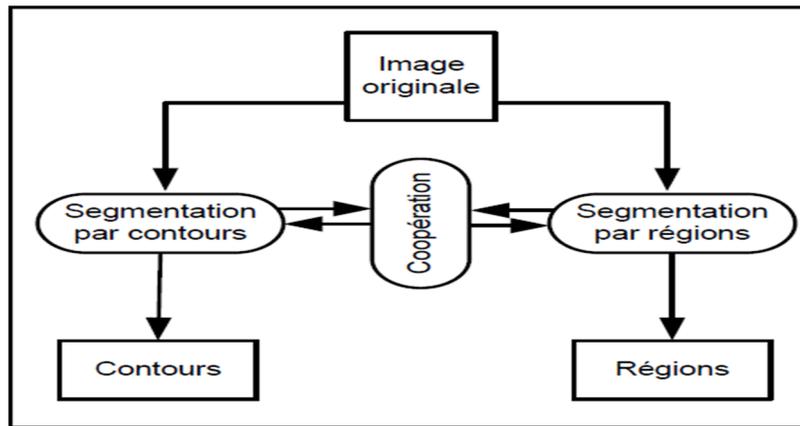


Figure 2.11 : synoptique de la coopération mutuelle

L'information échangée, sert à aider la technique de segmentation dans la prise de décision dans le cas de lacunes ou d'informations insuffisantes. La coopération permet de prendre des décisions plus sûres et plus fiables.

2.4 Introduction à la morphologie mathématique [18] [19]

La morphologie mathématique 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. En quelque sorte, chaque élément structurant fait apparaître l'objet sous une nouvelle forme. Fondamentalement, la morphologie mathématique a un caractère ensembliste et ceci, correspond à la démarche dont nous disposons pour analyser quantitativement une image.

2.4.1 Opérateurs de base et transformations premières de la morphologie mathématique

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

➤ Erosion

Soit X l'ensemble à éroder, et B un élément structurant de géométrie simple, par exemple un cercle. L'érosion consiste à faire déplacer B de telle sorte que son centre occupe toutes les

positions x de l'espace. Pour chaque position, on pose la question : est-ce que B est complètement inclus dans X ?

L'ensemble des positions x correspondant à une réponse positive forme le nouvel ensemble Y , appelé érodé de X par B , noté $X \ominus B$ (cf. Figure 2.12).

Cet ensemble satisfait l'équation 2.6 :

$$X \ominus B = \{x \text{ dans } X : B \subset X\} \quad 2.6$$

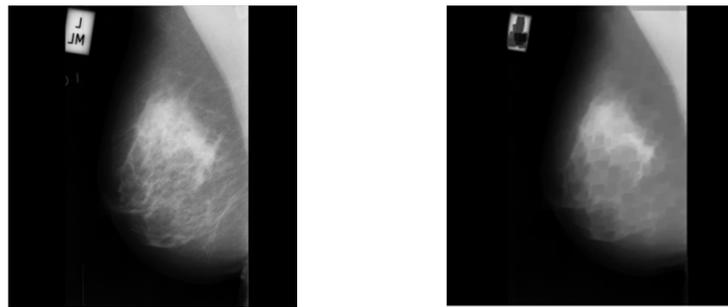


Image mammographique initiale Image mammographique érodée

Figure 2.12 : érosion d'une image mammographique

Après une érosion selon la figure ci-dessus;

- ✓ Les objets de taille inférieure à celle de l'élément structurant vont disparaître;
- ✓ Les autres seront "amputés" d'une partie correspondant à la taille de l'élément structurant;
- ✓ S'il existe des trous dans les objets, c'est à dire des "morceaux" de fond à l'intérieur des objets, ils seront accentués;
- ✓ Les objets reliés entre eux par un élément de taille inférieure à celle de l'élément structurant, vont être séparés.

➤ Dilatation

L'opération de dilatation se définit de manière analogue à l'érosion. En prenant le même élément structurant B , on pose pour chaque point x la question " Bx touche-t-il l'ensemble X ?", c'est à dire, y a-t-il une intersection non vide entre Bx et X ?



L'ensemble des points de l'image (cf. Figure 2.13) correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X, noté comme suit :

$$y = x \oplus B = \{ x \text{ dans } 1 : B \cap X \neq \emptyset \} \tag{2.7}$$

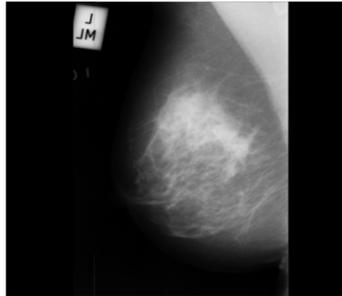


Image mammographique initiale

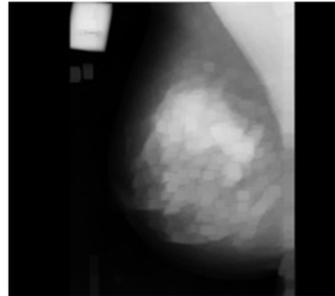


Image mammographique dilatée

Figure 2.13 : dilatation d'une image mammographique

➤ Ouverture

Puisque l'érosion et la dilatation sont des transformations itératives, il est possible d'effectuer sur un ensemble X une érosion, puis de dilater l'ensemble érodé par un même élément structurant B. Le résultat est nommé l'ouvert de X par B, noté comme suit :

$$O^B(X) = (X \ominus B) \oplus B \tag{2.8}$$

Après une ouverture, on ne retrouve pas l'ensemble de départ; L'ensemble ouvert est plus régulier et moins riche en détails que l'ensemble X initial (cf. Figure 2.14). La transformation par ouverture adoucit les contours, coupe les isthmes étroits, supprime les petites îles et les caps étroits.

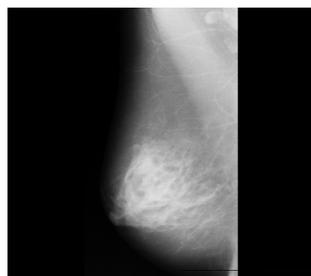


Image mammographique initiale

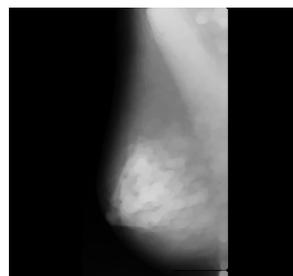


Image mammographique avec ouverture

Figure 2.14 : ouverture d'une image mammographique par un élément structurant circulaire

L'opération d'ouverture selon la figure ci-dessus, permet:

- ✓ de lisser les formes,
- ✓ d'éliminer les composantes connexes plus petites que B,
- ✓ de conserver souvent la taille et la forme, de ne pas conserver nécessairement la topologie.

➤ **Fermeture**

La fermeture est l'opération "inverse" de l'ouverture, c'est à dire que l'on applique tout d'abord une dilatation puis une érosion (toujours en gardant le même élément structurant). Le résultat est nommé le fermé de X par B, noté comme suit :

$$F^B(X) = (X \oplus B) \ominus B \quad \mathbf{2.9}$$

Un ensemble fermé, est également moins riche en détails que l'ensemble initial (cf. Figure 2.15). La transformation par fermeture bouche les canaux étroits, supprime les petits lacs et les golfes étroits.



Image mammographique initiale Image mammographique avec fermeture

Figure 2.15 : fermeture d'une image mammographique par un élément structurant circulaire

L'opération de fermeture suivant le résultat de la figure ci-dessus permet de:

- ✓ boucher les trous plus petits que B,
- ✓ conserver souvent la taille et la forme,
- ✓ ne pas conserver nécessairement la topologie,
- ✓ souder les formes proches en particulier.

2.4.2 Le gradient morphologique

C'est un algorithme intéressant permettant de mettre en évidence les contours d'objets dans une image. Le principe, est de matérialiser les fortes variations de niveaux de gris à l'aide de la dilatation et de l'érosion. Le gradient morphologique (cf. Figure 2.16) est défini comme étant la différence entre la dilatation d'une image par un élément structurant B et, son érosion par le même élément structurant. Pour un pixel donné, le gradient morphologique est donc la différence entre le maximum et le minimum des niveaux de gris des pixels traités par l'élément structurant B [2].

Ce gradient est noté ρ :

$$\rho_B(f) = \delta_B - \varepsilon_B \quad 2.10$$

On peut voir dans cette équation, que le gradient morphologique mesure la différence maximale de niveau de gris entre les pixels du voisinage défini par B.

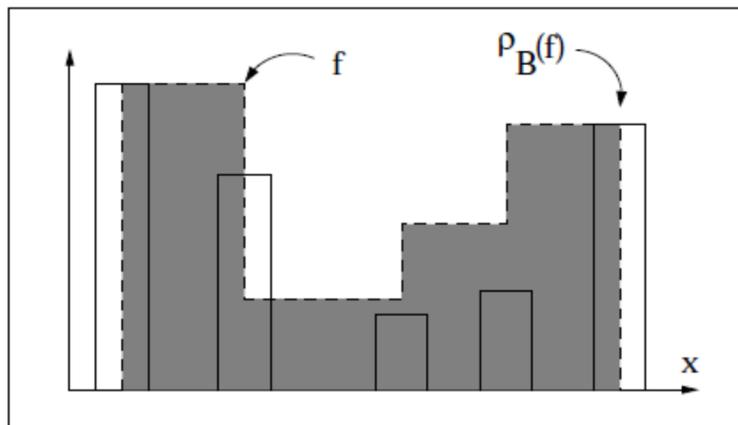


Figure 2.16: Gradient de Beucher d'une fonction f [2]

2.4.3 Transformation chapeau haut de forme [19][20]

Le chapeau haut de forme est une transformation qui permet, de retrouver les informations éliminées par l'ouverture morphologique ou, la fermeture morphologique. On définit la transformation du chapeau haut de forme, blanc comme la différence algébrique entre la fonction de départ f et l'ouvert $\gamma_B(f)$ (cf. Figure 2.17) tel que :

$$WTH_B(f) = f - \gamma_B(f) \quad 2.11$$

De même on définit, la transformation du chapeau haut de forme, noir (cf. Figure 2.18) comme :

$$BTH_B(f) = \phi_B(f) - f \tag{2.12}$$

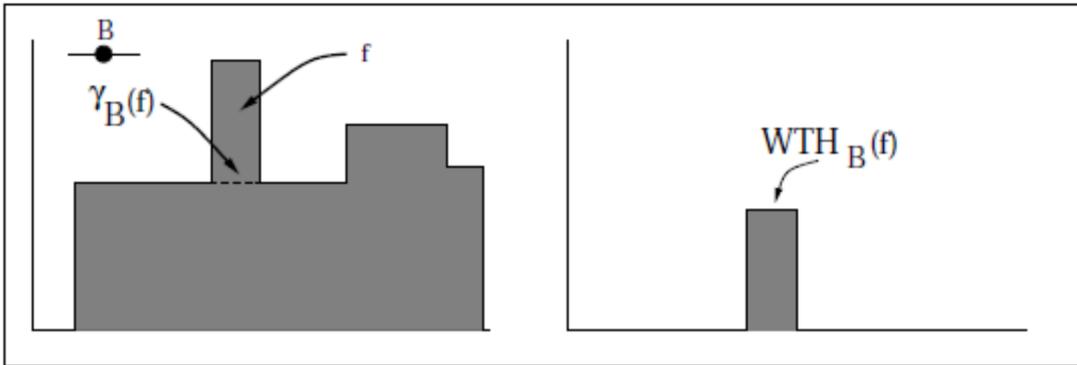


Figure 2.17 : chapeau haut de forme, blanc WTH d'une fonction numérique f par un Élément structurant B [19]

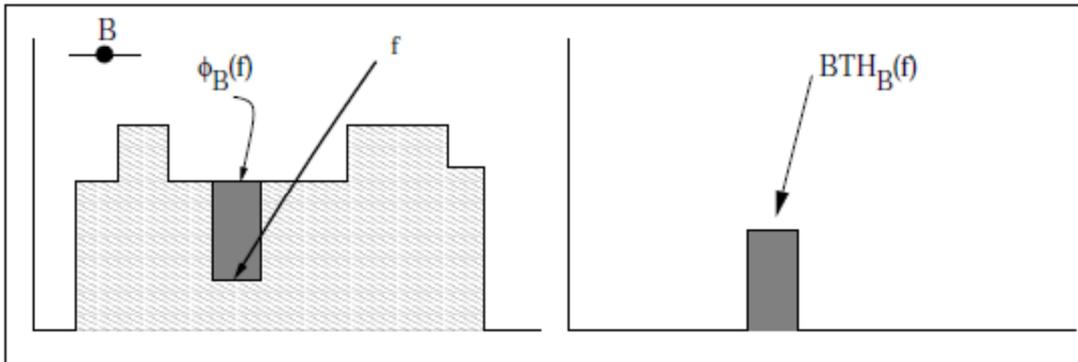


Figure 2.18 : chapeau haut de forme noir BTH d'une fonction numérique f par un élément Structurant B [19]

2.5 Aperçu superficiel des travaux en imagerie mammographique

Les attributs pour la détection des lésions mammographiques proposés dans la littérature, sont très variés et, on retrouve des adaptations de toutes les méthodes du traitement d'images mettant en œuvre, l'analyse multirésolution [21][22], la segmentation par croissance de régions adaptative [23][24], La détection par la morphologie mathématique



mise en exergue par la ligne de partage des eaux dans les travaux de Vachier [25], pour la détection des masses, etc...

L'équipe d'imagerie pour l'aide à la décision du laboratoire LATSI, s'intéresse à la détection de tumeurs mammaires, issues des images mammographiques et cytologiques et; consacre ses travaux, à l'analyse topologique et texturale pour la réalisation d'un système d'aide au diagnostic, suivant une classification connexionniste, par les arbres de décision ainsi que par les SVM [26][27][28][29]. L'objectif de ce travail, est de réaliser, une classification coopérative entre les SVM et le modèle du perceptron multicouche.

2.6 Méthodologie adoptée de détection des masses et calcifications

2.6.1 Cas des calcifications

Les deux principales caractéristiques des microcalcifications, sont leur petite taille et leur atténuation élevée. Une méthode de détection des microcalcifications basée sur l'opérateur chapeau haut de forme est étudiée dans le cas, de notre projet.

L'opérateur chapeau haut de forme avec seuillage (cf. Figure 2.19), est composé de trois étapes :

- l'ouverture de l'image avec un élément structurant adapté à la structure recherchée,
- la soustraction de l'image résultant de l'ouverture à l'image originale,
- L'application d'un seuil pour l'extraction des régions d'intérêt.

L'ouverture élimine toutes les structures plus petites que l'élément structurant et ne garde que les structures plus grandes. La soustraction de cette image à l'image originale, contient donc uniquement les petites structures qui ont été effacées par l'ouverture.

Finalement, le seuillage ne prend en compte que les structures présentant un certain contraste. Le seuillage se fait sur la somme des images du chapeau haut de forme, ce qui revient à donner plus d'importance aux petites structures.

2.6.2 Cas des masses

La détection de masses, se fait suivant la coopération séquentielle, entre la segmentation par analyse d'histogramme et, les contours actifs. L'information fournie par cette coopération permet d'éliminer les faux segments et de gagner en temps de traitement.

Par conséquent, la méthode de segmentation par analyse d'histogramme, va être guidée par les informations de la carte des contours, préalablement extraits. La carte des masses, est ainsi, obtenue.

Le système conçu, est donné par la figure 2.21 :

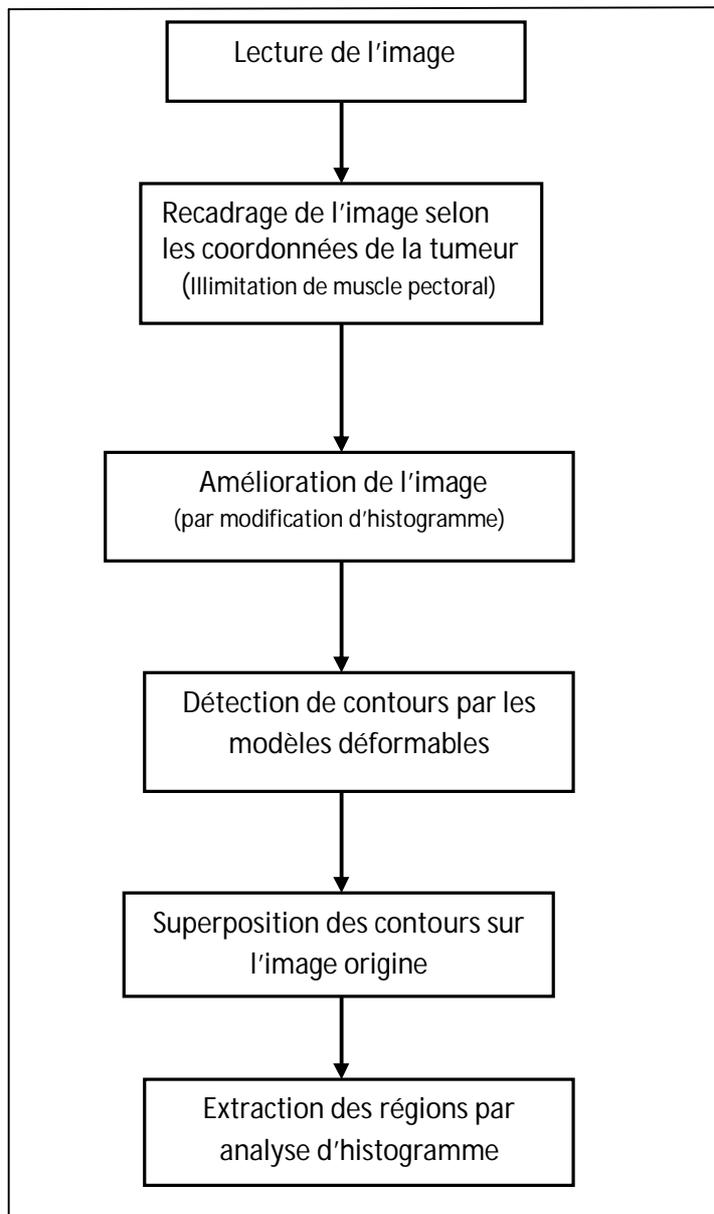
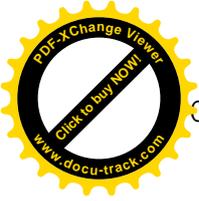


Figure 2.21 : système de détection des masses



2.7 Conclusion

Nous avons présenté dans ce chapitre, quelques techniques fondamentales, dans la segmentation d'image. Il ressort, que le domaine de segmentation est tellement riche et diversifié, qu'il est difficile de donner une étude exhaustive; celle-ci est déclinée en deux grandes familles de segmentation. Bien qu'elles aient toutes des avantages, elles présentent également plusieurs inconvénients. Dans ce contexte, nous nous sommes intéressés aux approches coopératives, entre la segmentation par analyse d'histogramme et la détection de contours par les modèles déformables; afin de, tirer avantage de la dualité des deux approches pour une meilleure segmentation. Ce chapitre a fait l'objet aussi, des principales méthodes de la morphologie mathématique notamment, la transformation chapeau haut de forme utilisée dans le cadre de ce travail, pour la détection des calcifications.

La segmentation d'image, est considérée comme l'étape fondamentale du processus d'analyse d'images mammographiques dédiés, à la détection et l'identification de la tumeur. L'adoption de la coopération entre les techniques de segmentation dans ces processus, permettrait de renforcer la qualité et la fiabilité des analyses et des décisions envisageables. Le chapitre suivant, est consacré à la caractérisation et morphologique des pathologies mammographiques, pour aborder la dernière phase de la détection des calcifications, qui se rapporte à la classification par la coopération entre les réseaux de neurones et, les machines à support vectoriel.



Chapitre 3

Caractérisation d'images

mammographiques et classification connexionniste.

3.1 Introduction

L'analyse de texture, regroupe un ensemble de techniques mathématiques, permettant de quantifier les différents niveaux de gris présents dans une image en termes d'intensité ou de rugosité et, leur distribution. Plusieurs méthodes d'analyse de texture existent et peuvent être classées en deux grandes catégories: les méthodes structurelles et les méthodes statistiques [30]. Les méthodes structurelles permettent de décrire la texture en définissant les primitives et les "règles" d'arrangement qui les relient. Les méthodes statistiques étudient quant à elles, les relations entre un pixel et son voisinage. Elles sont définies selon divers ordres: 1, 2, 3, ..., n (l'ordre d'un paramètre de texture pouvant être perçu comme étant le nombre de pixels pris en compte dans une itération lors de l'évaluation de ce paramètre). L'analyse de la topologie, repose sur la caractérisation de la géométrie, suivant une description spatiale des pixels.

Ce chapitre introduit un point très important de notre travail: le problème de l'extraction de caractéristiques. Par ce terme, nous entendons l'ensemble des méthodes permettant d'extraire des informations relatives à la texture et à la topologie, mais aussi et surtout au contenu structurel de l'image. Cette caractérisation nous dirige vers la classification connexionniste, étape finale de notre système



3.2 Analyse de texture

La notion de texture en traitement d'image, est d'autant plus délicate à aborder que les définitions que l'on en donne, sont multiples. Ces définitions, sont souvent liées à un aspect particulier mais sont rarement génériques et, la quantité importante d'approches utilisées pour l'analyse des textures, témoigne de l'absence d'une définition précise. L'objet de ce chapitre préliminaire, est tout d'abord de faire une synthèse des différentes interprétations, de la notion de texture. Bien que la notion de texture soit naturelle pour l'être humain, elle résiste depuis Longtemps à toute tentative de définition. On peut s'en approcher en disant qu'une texture est une zone de l'image qui présente certaines caractéristiques d'homogénéité qui la font apparaître comme une zone unique. Nous pouvons aussi la décrire comme étant un ensemble de primitives de taille et de forme variables, présentant une organisation spatiale particulière.

3.2.1 Définition de la texture [31]

Donner une définition précise et universelle de la notion de texture est un problème ardu qui n'a toujours pas été résolu (le besoin est pourtant réel, puisqu'une meilleure Compréhension de la notion de texture permettrait d'aborder son analyse de façon plus méthodique et plus concentrée; elle réduirait, sans doute, la profusion des approches trouvées actuellement dans la littérature). D'une part, le terme texture admet des acceptions différentes selon le domaine dans lequel il est utilisé (géologie, pétrologie, textile, agro-alimentaire,...), et, d'autre part, au sein d'un même domaine, et, particulièrement en traitement d'images, il n'existe pas de véritable consensus en ce qui concerne la définition d'une texture. Dire qu'une texture est une région d'une image présentant une organisation spatiale homogène des niveaux de luminance est correct mais très peu précis. Le manque crucial d'une définition exploitable de la notion de texture est, en partie, dû à la difficulté de comprendre et modéliser le système de perception visuelle humain. Une texture est une région de l'image, pour laquelle on peut définir une fenêtre De dimension minimale, telle qu'une observation au travers de celle-ci se traduit par une perception identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée. Pour une texture, il existe toujours une telle fenêtre qui correspond à la résolution de la texture. A la fois structurée et aléatoire, une texture peut être décrite spatialement ou statistiquement,

et, d'une manière générale, on qualifera une texture de fine ou grossière, plus ou moins homogène, rugueuse, plus ou moins régulière, dense, linéaire, isotrope ou directionnelle, tachetée, marbrée, etc.

Elle est peut-être détectée par sa primitive, le textuel (élément de texture), dans ce cas, la texture est une répétition des textuels.

3.2.2 Types de texture [31]

a. Texture périodique

Une texture périodique est formée de primitives arrangées d'une façon particulière, rangée. Le motif de base se répète alors, de manière régulière. On parle de texture ordonnée (cf. Figure 3.1).



Figure 3.1 : Exemples de textures périodiques

b. Textures aléatoire

Une texture aléatoire semble totalement désordonnée. Il est alors impossible d'isoler un motif de base (cf. Figure 3.2).



Figure 3.2 : Exemples de textures aléatoires



3.2.3 Introduction aux méthodes d'extraction des caractéristiques texturales [32]

Dans le traitement d'images, l'analyse de textures joue un rôle très important, pour distinguer les différentes régions d'une image, suivant l'information texturale.

Nous introduisons dans cette section, les principales méthodes d'extraction de paramètres permettant d'identifier les textures.

a. Méthodes structurelles

Ces méthodes, tiennent compte de l'information structurelle et contextuelle d'une forme et, sont particulièrement bien adaptées aux textures macroscopiques. Les étapes d'analyse, sont l'identification des éléments constitutifs et, la définition des règles de placement. Les deux structures les plus importantes, sont les structures de graphe et, les structures syntaxiques.

b. Méthodes statistiques

La texture est considérée comme la réalisation, d'un processus stochastique stationnaire. Des paramètres statistiques sont estimés, pour chaque pixel de l'image. Suivant la modalité des images à étudier, la signature la plus discriminante de la texture est à rechercher soit, dans des méthodes qui exploitent directement les propriétés statistiques de la texture (matrices de cooccurrences, matrice de longueurs de plages, matrice de voisinage, morphologie mathématique...), soit dans des méthodes qui exploitent les propriétés statistiques, à partir d'un plan transformé dans lequel on réécrit l'image de texture.

c. Méthodes fractales

Elles permettent de synthétiser des images très proches de la réalité. En analyse de texture, la dimension fractale, qui est une mesure du degré d'irrégularité d'un objet, décrit une certaine propriété de la texture. Le modèle fractal est basé essentiellement sur l'estimation, par des méthodes spatiales de la dimension fractale, de la surface représentant les niveaux de gris de l'image.

d. Méthodes fréquentielles

Ces méthodes sont basées sur la transformation de Fourier, afin d'établir des textures régulières et périodiques.

Dans le cas de notre travail, l'accent sera mis sur l'utilisation de la méthode statistique.

Nous expliciterons les paramètres extraits, suivant la matrice de cooccurrence.

3.1.4 Matrices de concurrence [30] [33]

Du fait de leur richesse en information de texture, les matrices de cooccurrences sont devenues les plus connues et, les plus utilisées pour extraire ces caractéristiques de textures. Elles estiment des propriétés des images relatives, à des statistiques de second ordre. Une matrice de cooccurrence, mesure la probabilité d'apparition des paires de valeurs de pixels situés à une certaine distance, dans l'image. Elle est basée sur le calcul de la probabilité $P(i, j, B, \theta)$ qui représente le nombre de fois où un pixel de niveau de couleur i , apparaît à une distance relative B , d'un pixel de niveau de couleur j et selon une orientation θ donnée.

Les directions angulaires θ classiquement utilisées, sont 0, 45, 90 et 135 degrés. Les relations de voisinage entre pixels, nécessaires au calcul des matrices, sont illustrées dans la figure (3.3). Par exemple, les plus proches voisins de 'x' selon la direction $\theta=135$ degrés, sont les pixels 4 et 8.

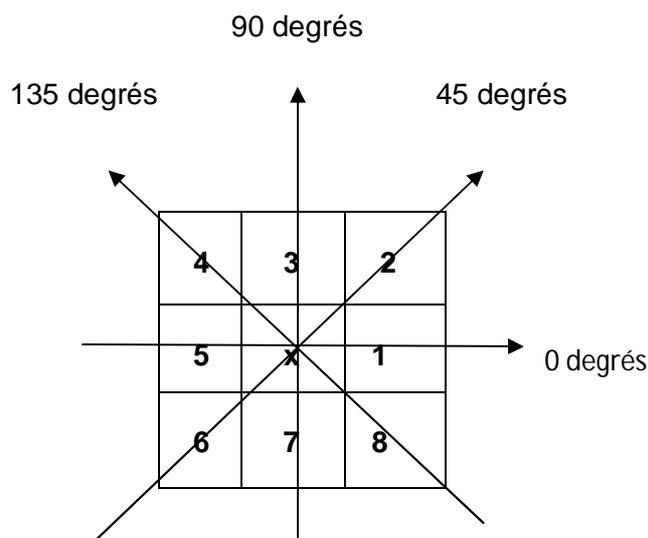


Figure 3.3: pixel x selon quatre directions



Les caractéristiques extraites à partir de ces matrices, contiennent des informations notamment sur l'homogénéité, les dépendances linéaires entre les niveaux de gris, le contraste et la complexité de l'image.

Nous définissons les 6 paramètres considérés, comme les plus pertinents, dans le cas de notre travail.

a. Energie

Ce paramètre (équation 3.1), mesure l'uniformité de la texture. Il atteint de fortes valeurs, lorsque la distribution des niveaux de gris est constante ou de forme périodique. Dans ce dernier cas, les valeurs élevées d'énergie, sont obtenues pour les matrices $P(\delta, \theta)$ lorsque (δ, θ) correspond à la période.

$$ENE = \sum_i \sum_j (P_{ij}(\delta, \theta)) \quad 3.1$$

b. Contraste

La valeur de ce paramètre (équation 3.2), est d'autant plus élevée, que la texture présente un fort contraste. Ce paramètre est fortement non corrélé, à l'énergie.

$$CST = \sum_i \sum_j ((i - j)^2 P_{ij}(\delta, \theta)) \quad 3.2$$

c. Entropie

Ce paramètre (équation 3.3), mesure le désordre dans l'image. Contrairement à l'énergie, l'entropie atteint de fortes valeurs, lorsque la texture est complètement aléatoire (sans structure apparente). Elle est fortement corrélée (par l'inverse) à, l'énergie.

$$ENT = \sum_i \sum_j (\log P_{ij}(\delta, \theta) P_{ij}(\delta, \theta)) \quad 3.3$$

**d. Variance**

La variance (équation 3.4), mesure l'hétérogénéité, de la texture. Elle augmente lorsque, les niveaux de gris diffèrent, de leur moyenne. La variance est indépendante du contraste.

$$\mathbf{VAR} = \sum_i \sum_j (i - u)^2 P_{ij}(\delta, \theta) \quad 3.4$$

e. Corrélation

La corrélation (équation 3.5), mesure la dépendance linéaire (relativement à (δ, θ)) des niveaux de gris de l'image. La corrélation n'est corrélée ni à l'énergie, ni à l'entropie.

$$\mathbf{COR} = \sum_i \sum_j \left[\frac{(i-u)(j-u)}{\sigma^2} P_{ij}(\delta, \theta) \right] \quad 3.5$$

f. Moment inverse

Le moment inverse (IDM) (équation 3.6), mesure l'homogénéité de l'image. Ce paramètre est corrélé à une combinaison linéaire de l'énergie et du contraste.

$$\mathbf{IDM} = \sum_i \sum_j \left(\frac{P_{ij}(\delta, \theta)}{1 + (i-j)^2} \right) \quad 3.6$$

3.3 Analyse de la forme [34][35]36]

Dans la littérature, des dizaines de paramètres de forme, permettent de caractériser la morphologie des particules. Une des difficultés de l'analyse morphologique, réside dans le choix du paramètre de forme le plus adapté à la nature des particules.

Les caractéristiques que nous avons choisies de calculer, ce sont celles qui nous paraissent les plus pertinentes, pour modéliser numériquement les régions d'intérêt mammaires. Elles permettent ainsi, d'identifier les zones cancéreuses, en se basant sur les descripteurs morphologiques.



3.3.1 Compacité

La compacité est une mesure simple d'une forme. Elle est donnée par la relation 3.7, telle que:

$$\text{Comp} = \frac{P^2}{A} \tag{3.7}$$

Où : P et A respectivement sont le périmètre et la surface de la forme de la région d'intérêt.

Une normalisation est effectuée suivant la relation:

$$\text{Compacité} = \frac{4\pi s(r)}{p^2(r)} \tag{3.8}$$

3.3.2 Excentricité

L'excentricité (équation 3.9) est le rapport entre la longueur du grand axe et, celle du petit axe. L'excentricité est calculée à partir du moment d'inertie.

$$\text{Excentricité} = \frac{M_{0.2} + M_{2.0} - \sqrt{(M_{0.2} + M_{2.0})^2 + 4(M_{1.1})^2}}{M_{0.2} + M_{2.0} + \sqrt{(M_{0.2} + M_{2.0})^2 + 4(M_{1.1})^2}} \tag{3.9}$$

3.3.3 Convexité et enveloppe convexe

La convexité représente le rapport entre la surface de la région et celle, de son enveloppe convexe cf. Figure 3.4).

L'enveloppe convexe est le plus polygone convexe, qui englobe tous les points. Ainsi, une région est fortement convexe, si la valeur de convexité est égal à 1. A l'inverse, une région est faiblement convexe, si la valeur de convexité inférieure à 1.

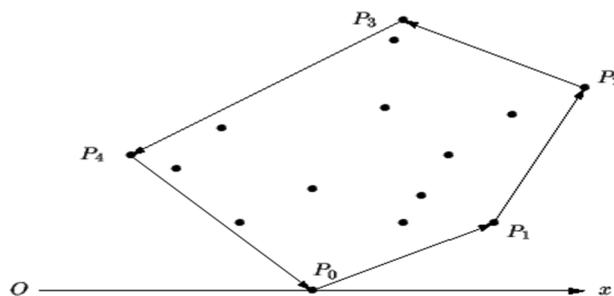


Figure 3.4 : enveloppe convexe



3.3.4 Moments invariants de Hu

Ces moments [34][36] permettent de décrire, la forme à l'aide de propriété statiques. Ils sont simples à manipuler et, sont robustes aux transformations géométriques; comme la translation, la rotation et le changement d'échelle.

Plusieurs techniques ont été développées, pour la caractérisation et la représentation d'objet par ces moments. Hu [36] a défini sept descripteurs invariants suivant ci-dessous:

$$\begin{aligned}
 \Phi_1 &= \eta_{20} + \eta_{02} \\
 \Phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11} \\
 \Phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (\eta_{30} - 3\eta_{21})^2 \\
 \Phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
 \Phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
 & \quad (3\eta_{21} - \eta_{03})(\eta_{23} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 \Phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
 \Phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]
 \end{aligned} \tag{3.10}$$

Ces descripteurs sont engendrés des moments;

$$\eta_{pq} = \frac{m'_{pq}}{m'^n_{00}} \tag{3.11}$$

où $n = (p + q) / 2$.

Avec m'_{pq} , représentant les moments centraux :

$$m'_{pq} = \sum_x \sum_y (x - x_g)^p (y - y_g)^q f(x, y) \tag{3.12}$$

Le centre de (x_g, y_g) représente le centre de gravité de l'image dont les coordonnées, sont définies par:

$$x_g = \frac{m_{10}}{m_{00}}, \quad y_g = \frac{m_{01}}{m_{00}} \tag{3.13}$$

Nous considérons dans le cadre de notre projet, la moyenne des descripteurs invariants.



3.4 Introduction à la classification d'image

Les méthodes de classification [35] [37], permettent de regrouper des objets, en classes d'objets plus homogènes. Les objets regroupés, ont des caractéristiques communes. Ils sont similaires mais, se distinguent clairement des objets, des autres classes.

La mise en œuvre d'une procédure de classification, ayant pour objectif de classifier automatiquement des objets, comporte généralement deux phases fondamentales:

- une phase d'apprentissage dont le but, est de déterminer un espace de représentation des données et, de rechercher les paramètres discriminants capables, de caractériser chaque classe d'objets ;
- une phase de reconnaissance au cours de laquelle, on attribue à une classe, chacun Des objets inconnus, dans l'espace de représentation déterminé durant l'apprentissage.

La classification peut être supervisée ou non supervisée.

Les méthodes de classification, ont pour but d'identifier les classes auxquelles appartiennent des objets, à partir de certains traits descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier, au problème de la prise de décision automatisée. La procédure de classification sera extraite, automatiquement à partir d'un ensemble d'exemples. Un exemple, consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification. Il s'agit à cet effet, d'extraire une règle générale à partir des données observées. La procédure générée, devra classifier correctement les exemples de l'échantillon et, avoir un bon pouvoir prédictif, pour classifier correctement de nouvelles descriptions.

Les méthodes utilisées pour la classification sont nombreuses, citons: la classification par les réseaux de neurones, les machines à supports vecteurs (SVM), etc.

Nous étudions dans ce mémoire la classification connexionniste, qui montre de bonnes performances, dans la résolution de problèmes réels, tels que; la catégorisation de textes ou le diagnostic médical et, ce même sur des ensembles de données de très grandes dimensions. Dans le cas de notre projet, nous considérons l'étude de la classification supervisée.



La classification supervisée, demande à l'utilisateur d'instruire le système, en désignant des zones de l'image, comme étant des échantillons représentatifs des classes à extraire. La classification est donc, précédée d'un apprentissage. Pour instruire le système, il faut avoir une bonne connaissance du terrain observé. Pour cela, il faut disposer d'une vérité terrain. C'est un outil pour réaliser un bon apprentissage, ainsi que pour valider une classification. Il est suffisant de disposer d'une vérité terrain sur une portion de l'image où les types de chaque région, sont représentés.

La classification supervisée, s'organise en quatre étapes:

- **espace des états** : on extrait de l'image, des critères jugés suffisamment discriminants pour la classification. Cet ensemble définit un espace, dont la dimension est égale au nombre de critères extraits;
- **zones d'apprentissage** : on définit sur l'image des zones d'apprentissage, qui sont des régions de l'image que l'on juge représentatives d'une certaine classe;
- **apprentissage** : il consiste, pour une classe donnée, à rassembler l'ensemble des pixels situés dans les zones d'apprentissage et, d'étudier la répartition de leurs vecteurs associés dans l'espace des états.
- **classifieur**: il utilise l'information issue de l'apprentissage, pour attribuer à chaque point, ou à chaque région de l'image, une catégorie ou une classe.

3.4.1 Apprentissage [37]

L'objectif de l'apprentissage à partir d'exemples étiquetés, appelé apprentissage supervisé, est de construire une fonction qui permet, d'approcher au mieux une fonction inconnue qui génère des données aléatoires. Un système d'apprentissage à partir d'exemples, est composé de trois modules principaux:

- un générateur qui génère des données aléatoires appelées les vecteurs d'entrée;
- un superviseur, qui associe pour chaque vecteur d'entrée x une sortie y (la classe) suivant une distribution de probabilité également inconnue $P(x, y)$;



- une machine d'apprentissage qui permet d'implémenter une famille de fonctions $f_{\alpha}(x); \alpha \in A$, où A est un ensemble de paramètres.

La figure 3.5 représente ces trois modules:

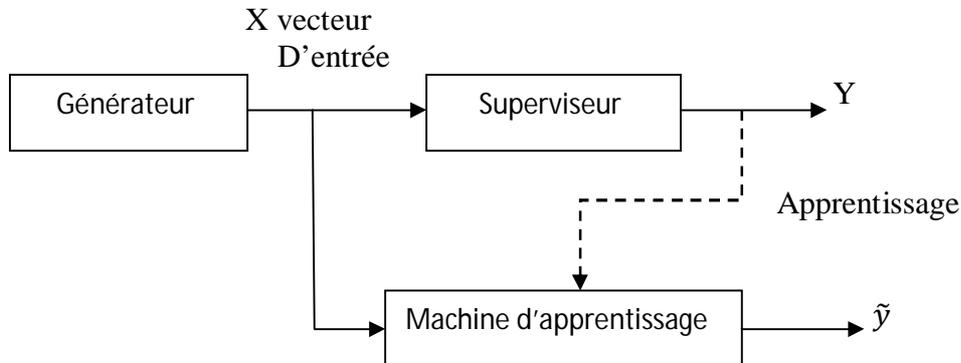


Figure 3.5: modules d'un système d'apprentissage

3.4.2 Apprentissage supervisé [37]

Soit \mathcal{O} un ensemble d'individus décrit, par un nombre fixe n de caractéristiques. Prenons un sous-ensemble \mathcal{S} de \mathcal{O} . Etant donnée les paramètres d'un objet, de \mathcal{O}/\mathcal{S} on veut pouvoir estimer, le mieux possible la classe de cet objet (**-1 ou +1**) avec la seule connaissance de \mathcal{S} . \mathcal{S} est l'ensemble d'apprentissage (training set) et \mathcal{O}/\mathcal{S} est, l'ensemble de test. On représente ainsi, \mathcal{S} par un ensemble de m couples (x, y) où x est un point de \mathbb{R}^n , qui représente les paramètres et $y_i = \pm 1$ représente la classe de x_i . Les données (x, y) sont issues, d'une distribution de probabilité inconnue $P(x, y)$. Les données (x, y) sont supposées indépendamment identiquement distribuées. Le but de l'apprentissage, est de trouver une estimation de l'application $x_i \rightarrow y_i$.

3.5 Réseaux de Neurones [38][39]

On considère généralement qu'un réseau de neurones, est constitué d'un grand ensemble d'unités (ou neurones), ayant chacune une petite mémoire locale. Ces unités, sont reliées par des canaux de communication (les connexions, aussi appelées synapses d'après le terme biologique correspondant), qui transportent des données numériques. Les unités, peuvent uniquement agir sur leurs données locales et, sur les entrées qu'elles reçoivent par leurs connexions.

3.5.1 Neurones biologiques

Le cerveau est composé de plusieurs cellules que l'on peut diviser en deux grands groupes : d'un côté, les neurones et de l'autre, l'ensemble des cellules de support que l'on nomme 'neuroglia'. Il n'y a que les neurones qui soient des éléments actifs dans le cerveau, les cellules gliales ne servent qu'à supporter physiquement les neurones. Elles agissent aussi comme isolateur électrique, elles aident au métabolisme et, fournissent des protéines aux neurones (cf. Figure 3.6).

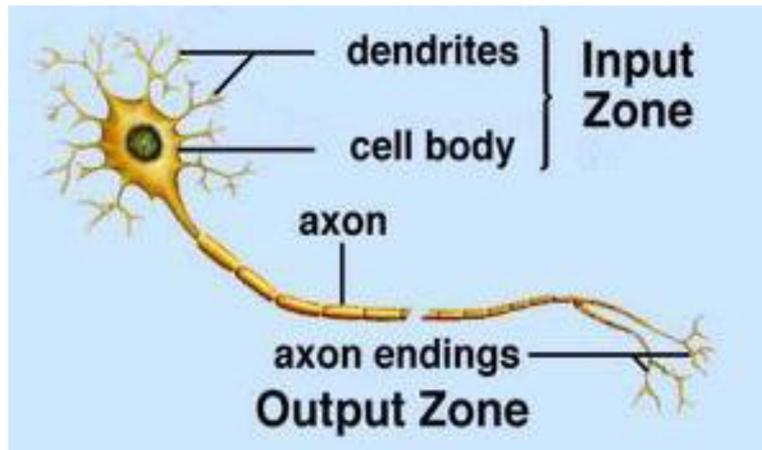


Figure 3.6 : neurone biologique

3.5.2 Neurones artificiels

Un réseau de neurones artificiel, est un ensemble de neurones interconnectés. Le neurone est un élément contenant plusieurs entrées et, une sortie. Chaque entrée x_i est pondérée par un poids w_{ki} , avant d'être sommé pour produire v_k , k étant un indice, indiquant le numéro du neurone. Le résultat v_k est soumis à, une fonction de transfert $\varphi(.)$ non linéaire (cf. Figure 3.7).

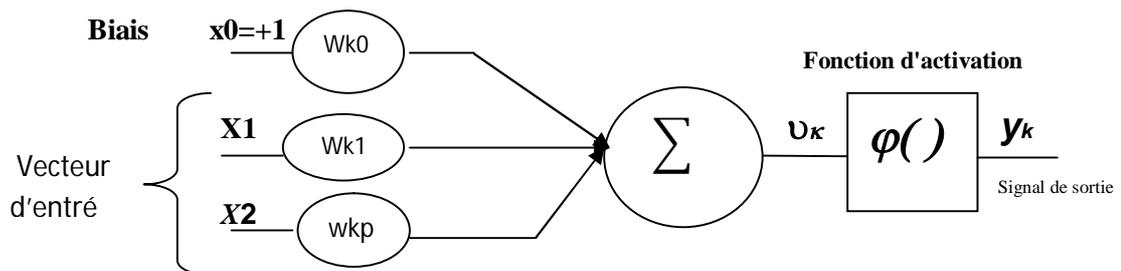


Figure 3.7 : représentation schématique d'un neurone artificiel



3.5.3 Principe de fonctionnement [37]

L'information ne se transmet que dans un seul sens: des dendrites vers l'axone. Le neurone va donc recevoir des informations, venant d'autres neurones, grâce à ses dendrites. Il va ensuite y avoir une sommation, au niveau du corps cellulaire; de toutes ces informations et via, un signal électrique. Le résultat de l'analyse, va transiter le long de l'axone, jusqu'aux terminaisons synaptiques. Il va y avoir, libération des neurotransmetteurs (médiateurs chimiques) dans la fente synaptique. Le signal électrique, ne pouvant pas passer la synapse, les neurotransmetteurs permettent donc le passage des informations, d'un neurone à un autre. Les neurotransmetteurs excitent ou inhibent, le neurone suivant et, peuvent ainsi générer ou interdire, la propagation d'un nouvel influx nerveux. Les synapses possèdent une sorte de «mémoire» qui, leur permet d'ajuster leur fonctionnement. En fonction de leur «histoire», c'est-à-dire de leur activation répétée ou non, entre deux neurones, les connexions synaptiques vont se modifier: c'est le mécanisme d'apprentissage.

3.5.4 Calcul des poids synaptiques [38]

La rétro-propagation est une méthode de calcul des poids (aussi appelés poids synaptiques du nom des synapses, terme désignant la connexion biologique entre deux neurones) pour un réseau à apprentissage supervisé; qui consiste, à minimiser l'erreur quadratique de la sortie (somme des carrés de l'erreur de chaque composante entre la sortie réelle et la sortie désirée).

D'autres méthodes de modification des poids sont plus locales, chaque neurone modifie ses poids en fonction de l'activation ou non des neurones proches.

3.5.5 Les grands types de réseaux [39]

Il existe plusieurs types de réseaux de neurones, nous étudions les plus connus et nous décrivons davantage, le perceptron multicouche.

a. Réseaux de Hopfield

Ces réseaux sont des réseaux récurrents, un peu plus complexes que les perceptrons multicouches. Chaque cellule est connectée à toutes les autres et, les changements de

valeurs de cellules s'enchainent en cascade jusqu'à un état stable. Ces réseaux sont bien adaptés à la reconnaissance de formes.

b. Réseaux de Kohonen

Ces réseaux, peuvent détecter des régularités et des corrélations, dans les données d'entrées et peuvent adapter leurs réponses, en fonction de ces entrées.

Les neurones de ces réseaux, peuvent reconnaître des groupes dans les vecteurs d'entrée. Les cartes auto organisatrices, peuvent apprendre à reconnaître des groupes de vecteurs similaires, ce qui entraîne que les neurones physiquement proches les uns des autres dans la couche de neurones, à répondre à ces vecteurs d'entrée similaires.

C. Perceptrons [39] [40]

1. Perceptron simple

Le perceptron simple, permet de classifier des éléments compris dans deux catégories différentes; à condition que ces deux catégories, soient linéairement séparables. Par la suite, il est également possible de prédire, à quelle catégorie appartient un élément.

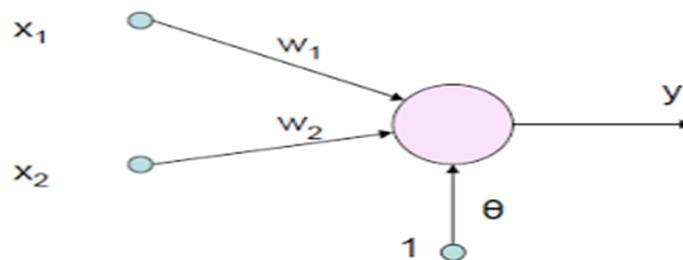


Figure 3.8 : perceptron simple

2. Perceptron multicouche (PMC)

Malgré son nom quelque peu barbare, le perceptron multicouche, est sans doute le plus simple et le plus connu des réseaux de neurones. Il s'agit d'un classifieur linéaire de type réseau neuronal formel organisé en plusieurs couches, au sein desquelles une information, circule de la couche d'entrée vers la couche de sortie, uniquement. Chaque couche, est constituée d'un nombre variable de neurones, les neurones de la couche de



sortie correspondant toujours aux sorties du système. La structure du (PMC) est relativement simple (cf. Figure 3.9):

- ✓ **Les nœuds d'entrée** : la première couche, est appelée couche d'entrée. Elle recevra les données source, que l'on veut utiliser pour l'analyse. Dans le cas de l'aide au diagnostic médical, cette couche recevra les symptômes. Sa taille est donc directement déterminée, par le nombre de variables d'entrées.

- ✓ **Les nœuds cachés** : La seconde couche est une couche *cachée*, en ce sens qu'elle n'a qu'une utilité intrinsèque pour le réseau de neurones et, n'a pas de contact direct avec l'*extérieur*. Les fonctions d'activations, sont en général non linéaires sur cette couche, mais il n'y a pas de règle à respecter. Le choix de sa taille n'est pas implicite et doit être, ajusté. En général, on peut commencer par une taille moyenne des couches d'entrée et de sortie mais ce n'est pas toujours, le meilleur choix. Il sera souvent préférable, pour obtenir de bon résultats, d'essayer le plus de tailles possibles.

- ✓ **Les nœuds de sortie** : la troisième couche, est appelée couche de sortie. Elle donne le résultat obtenu, après compilation par le réseau des données de l'entrée, dans la première couche. Dans le cas de l'aide au diagnostic médical, cette couche indique le type de pathologie. La taille de cette couche, est directement déterminée par le nombre de variables que l'on veut, en sortie.

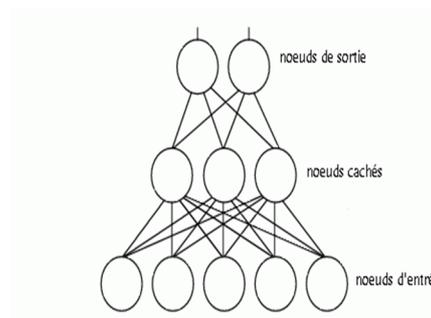


Figure 3.9: perceptron multicouche



3.5.6 Fonction d'activation [39]

Comme indiqué précédemment, un perceptron multicouche (MLP) est un réseau de neurones avec une architecture « **feed-forward** » et des connexions unidirectionnelles complètes entre les différentes couches. Toutefois, ce n'est pas suffisant pour définir les propriétés d'un réseau. Outre les architectures de réseaux, les neurones d'un réseau possèdent des fonctions d'activation, qui vont transformer les signaux émis par les neurones de la couche précédente à l'aide, d'une fonction mathématique. Cette fonction f , est la fonction d'activation elle-même, qui, peut profondément influencer sur la performance du réseau. Il est donc important de bien choisir ce type de fonction d'activation des neurones, dans un réseau de neurones. Les neurones d'entrée, ne possèdent généralement aucune fonction d'activation.

La fonction f , peut être paramétrée de façon quelconque, mais deux types de paramétrage sont fréquemment utilisés:

- ✓ Les paramètres sont attachés aux entrées du neurone; la sortie est une fonction non linéaire d'une combinaison des entrées, pondérées par les paramètres.
- ✓ Les paramètres sont attachés à la non-linéarité du neurone et ils interviennent directement, dans la fonction f .

3.5.7 Avantages et inconvénients

Parmi les avantages des réseaux, nous citons :

- ✓ précision du classifieur (si bien paramétré);
- ✓ apprentissage automatique des poids;
- ✓ possibilité de faire le parallélisme (les éléments de chaque couche peuvent fonctionner en parallèle);
- ✓ résistance aux pannes (si un neurone ne fonctionne plus, le réseau ne se perturbe pas).

Les inconvénients se résument suivant;

- ✓ la complexité de l'architecture du réseau;



- ✓ paramètres difficiles à interpréter (boite noire);
- ✓ difficulté du paramétrage, surtout pour le nombre de neurones, dans la couche cachée.

3.6 Algorithme génétique

3.6.1 Définition

L'algorithme génétique (AG) est un algorithme de recherche basé sur les mécanismes de la sélection naturelle et de la génétique. Il combine une stratégie de « survie des plus Forts » avec un échange d'information aléatoire mais structuré. Pour un problème pour lequel une solution est inconnue, un ensemble de solutions possibles est créé aléatoirement. On appelle cet ensemble la population. Les caractéristiques (ou variables à déterminer) sont alors utilisées dans des séquences de gènes qui seront combinées avec d'autres gènes pour former des chromosomes et par conséquent des individus. Chaque solution est associée à un individu, et cet individu est évalué et classifié selon sa ressemblance avec la meilleure, mais encore inconnue, solution au problème. Il peut être démontré qu'en utilisant un processus de sélection naturelle inspiré de Darwin, cette méthode convergera graduellement à une solution.

L'algorithme génétique de base comporte trois opérations simples qui ne sont pas plus compliquées que des opérations algébriques :

- Sélection
- croisement
- Mutation

3.6.2 Opérateur génétique

a. Sélection : processus où les individus sont copiés selon la valeur de leur fonction objective f . On peut écrire la fonction f comme une mesure de profit, utilité ou qualité que l'on veut maximiser (minimiser). Si on copie des individus selon leur valeur f , ceci implique que les individus ayant des valeurs plus élevées ont une plus grande probabilité de contribuer des rejetons à la prochaine génération. Ceci correspond à une version artificielle de la "survie des plus forts" (*survival of the fittest*) de Darwin.

L'implantation de la sélection peut se faire de plusieurs façons. La plus facile est peut-être la roue de roulette biaisée, où la probabilité de reproduction d'un individu dépend de sa valeur par rapport au total des valeurs de la population



b. Croisement : processus où de nouveaux individus sont formés à partir de parents. Ces nouveaux individus, les rejetons, sont formés en effectuant un croisement entre deux parents. On choisit une position aléatoire k entre $[1, L-1]$ où " L " est la longueur de l'individu. Le croisement se fait en échangeant les bits de la position $k+1$ à L .

c. Mutation : processus aléatoire où un bit change de valeur. Ce processus joue un rôle secondaire dans l'algorithme génétique, mais il est quand même important. La mutation assure qu'aucun point dans l'espace de recherche n'a une probabilité nulle d'être atteint.

3.6.3 Principe de fonctionnement

Un algorithme génétique recherche les extrêmes d'une fonction définie sur un espace de données appelé population initiale. Par analogie avec la génétique, chaque individu de cette population est un chromosome et chaque caractéristique de l'individu est un gène.

Dans un cas simple, un gène sera représenté par un bit (0 ou 1), un chromosome par une chaîne de bits et un individu par un ensemble de chaînes de bits.

Pour commencer, l'AG génère aléatoirement une population initiale (comme solutions possibles). Il opère, ensuite, à un croisement des meilleurs chromosomes (les meilleurs sont choisis par une fonction d'évaluation propre au problème à résoudre). Ce croisement (hybridation), consiste en l'échange d'un certain nombre de bits (gènes) entre les deux parents. Les meilleurs enfants obtenus seront croisés, à leur tour, pour obtenir encore une meilleure génération.

L'algorithme (cf. Figure 3.10) crée des mutations (change la valeur de quelques bits aléatoirement), pour bien imiter le processus naturel. On répète ces étapes, jusqu'à ce qu'à ce qu'un enfant soit estimé proche de la solution recherchée.

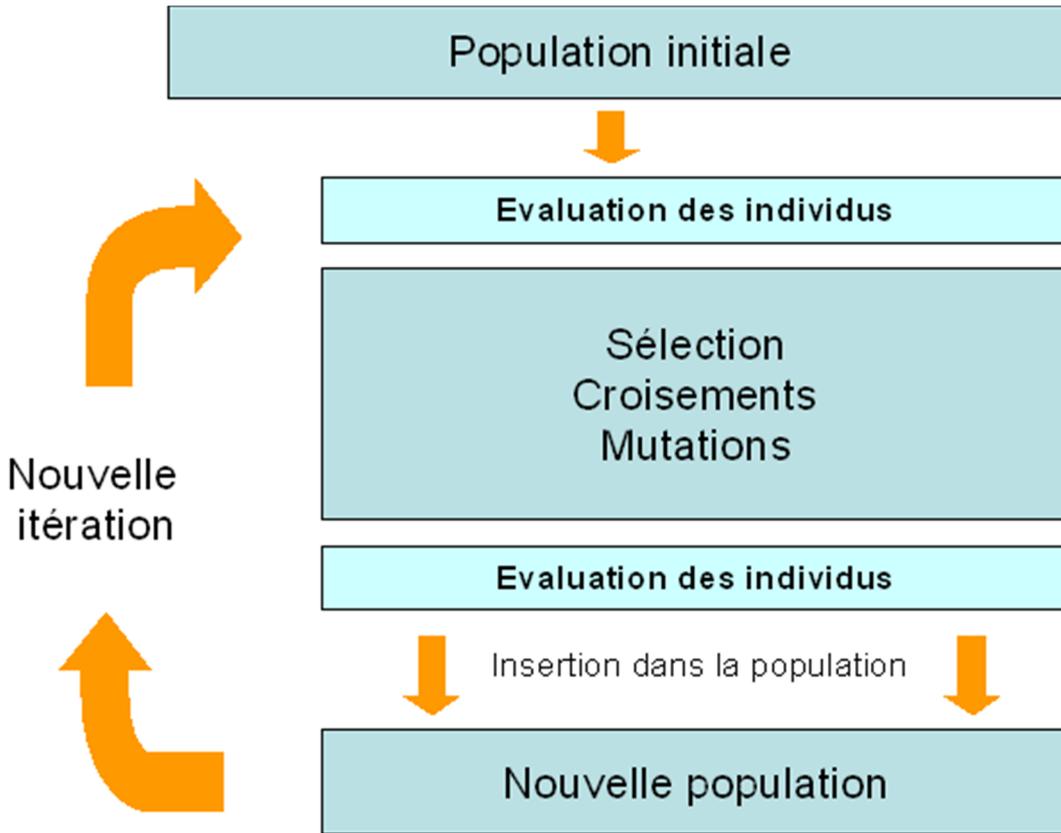


Figure 3.10 : principe de fonctionnement d'AG

3.6.4 avantage d'AG

Contrairement à la recherche opérationnelle, l'AG n'exige aucune connaissance de la manière dont on peut, résoudre le problème. Il est seulement nécessaire, de pouvoir évaluer la qualité de la solution. Également, dans le cas de recherche d'optimum de fonctions analytiques, on n'a besoin ni, de dérivabilité ni de continuité

Un algorithme génétique, nous donne une grande liberté, dans le paramétrage et dans l'implémentation, des différents traitements. Libre à nous, de modifier tel ou tel, paramètre si les solutions obtenues ne nous conviennent pas.

Les algorithmes génétiques, ont l'énorme avantage de pouvoir être appliqués, dans un grand nombre de domaines de recherche de solution; lorsqu'il n'est pas nécessaire d'avoir la solution optimale, qui prendrait par exemple trop de temps et, de ressources pour être calculée (ou tout simplement si personne n'est capable, de la trouver de manière théorique).

Dans le cas de notre étude, l'AG a l'un avantage de rechercher la solution optimale pour trouver les meilleurs paramètres du réseau de neurones



3.7 Classification en SVNN

On sait que ce qui caractérise un réseau de neurones, ce sont ses poids (W_i) et ses biais (B_i), c'est cela qui relie toutes les entrées à tout le neurone, jusqu'à la couche de sortie.

Dans un RN classique, on intègre les entrées dans la couche primaire et, l'apprentissage se fait avec un certain nombre d'itérations et de tests. L'arrêt s'effectue, lorsque l'erreur quadratique moyenne est basse et, donc le taux de réussite augmente. Cependant, le défaut est, à chaque fois, qu'on recommence l'entraînement, on a un résultat différent; c'est donc un processus aléatoire, qui ne donne pas toujours un très bon résultat, surtout quand les classes sont difficilement séparables.

Avec la classification SVNN (support vector neural network), lorsqu'on spécifie le nombre de neurones, le réseau est automatiquement créé. Les biais et les poids sont initialisés aléatoirement.

À la première itération: on fait une comparaison des entrées avec les sorties et, on voit si les biais et les poids sont ajustés. Évidemment, comme ils ont été initialisés aléatoirement, ils sont obligatoirement faux, donc, c'est là que les AG prennent leur importance, avec l'évaluation de ses individus (W_i , B_i) on peut effectuer l'opération de sélection, croisement, mutation et, une nouvelle fois, l'évaluation. À partir de là, une nouvelle population est créée et, à nouveau évaluée. Le réseau est de nouveau, évolué suivant les opérations de sélection, croisement et, mutation et, ainsi de suite, avec des dizaines, voire des centaines d'itérations.

À la fin, on obtient les poids et biais optimaux, ce qui nous donne un réseau de neurone proche de la perfection, sans avoir à faire une multitude d'entraînements, sans intérêt.

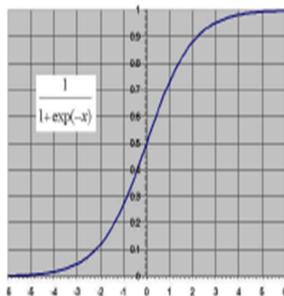
3.7.1 Création de la base d'apprentissage

Lors de la phase d'apprentissage, (cf. Figure 3.10) le neurone modifie ses poids. Nous allons maintenant expliquer comment cette modification s'effectue en amont du neurone, par la règle de rétropropagation du gradient et la règle de Hebb [39]. L'algorithme de rétropropagation du gradient, consiste à déterminer l'erreur effectuée par chaque neurone puis à modifier la valeur des poids pour minimiser cette erreur. Il faut effectuer des rétropropagations jusqu'à ce que l'erreur quadratique moyenne devienne inférieure à un

certain seuil. L'erreur quadratique est définie comme la somme sur les sorties des carrés des différences entre l'entrée obtenue et l'entrée désirée soit :

$$E = \sum (S_i - Y_i)^2 \tag{3.15}$$

Le réseau est représenté par 3 couches de neurones, chaque neurone d'une couche étant relié à tous les neurones de la couche suivante par une liaison pondérée. Chaque neurone intègre les données grâce à une fonction non linéaire qui permet l'apprentissage de classes non linéairement séparables [14]. On utilise ici la fonction sigmoïde suivante :



$$s(x) = \frac{1}{1 + e^{-x}} \tag{3.16}$$

$$ds(x) = s(x)(1 - s(x))$$

Figure 3.10: Représentation de la fonction sigmoïde [39]

On effectue tout d'abord une propagation de l'entrée à travers le réseau. Considérons S le vecteur de sortie obtenu après la propagation de l'entrée X à travers le réseau, et Y le vecteur des sorties que l'on aurait voulu obtenir. Pour la couche de sortie considérons le neurone i ; on obtient donc l'erreur pour ce neurone : $S[i]-Y[i]$

On appelle p la taille de la couche cachée. Le neurone i reçoit alors les p sorties des neurones de la couche cachée modulées de leur poids respectifs. Ce sont ces poids qu'il nous faut modifier. Pour relativiser l'importance d'un poids par rapport à un autre, dans l'erreur totale il faut moduler sa modification par la sortie du neurone auquel il est rattaché [35]. Il faut maintenant utiliser le gradient de l'erreur afin de pouvoir à terme faire converger la valeur du poids. On obtient alors la formule suivante :

La règle de modification des poids pour un exemple k :

$$W_{ij}(k) = W_{ij}(k - 1) - pas \times D_i \times O_j \times ds(I_i) \tag{3.17}$$

$$D_i = (S_i - Y_i) \text{ Pour la couche de sortie,}$$



Avec :

W_{ij} : le poids entre le neurone i d'une couche et le neurone j de la couche précédente

pas : pas de l'apprentissage

O_j : sortie du neurone j

I_i : entrée du neurone i : $I_i = \sum_j W_{ij} \cdot O_j$

S_i : Sortie obtenue,

Y_i : Sortie désirée,

ds : Fonction sigmoïde dérivée.

Pour la couche cachée, déterminer l'erreur commise par chaque neurone n'est pas instantanée ; puisque l'on ne dispose pas de la valeur théorique qu'il devrait y avoir. La méthode, consiste à remarquer que l'erreur effectuée par un neurone de la couche cachée est la somme des erreurs des neurones de la couche de sortie modulée par les poids respectifs des liaisons entre les deux couches. Ainsi l'erreur pour le neurone i , de la couche cachée devient [38]:

$$\sum_h W_{hi} \cdot (S[h] - Y[h]) \quad h \text{ appartenant à la couche de sortie.}$$

On obtient alors la formule de modification des poids suivante :

$$W_{ij}(k) = W_{ij}(k-1) - \text{pas} \times D_i \times O_j \times ds(I_i) \quad 3.18$$

$$D_i = \sum_h D_h \cdot W_{hi} \quad \text{pour la couche cachée}$$

L'approche des SVM un contre un, permet dans le cadre de l'estimation du classifieur du perceptron, sous l'hypothèse que la frontière, séparant une catégorie d'une autre, peut être moins complexe que celle, séparant cette même catégorie de toutes les autres [42].



1- Déclaration des paramètres qui caractérisent le réseau :

K, pas, seuil.

2- Initialisation du vecteur des poids par des valeurs aléatoires :

W_{ij} est le poids entre le neurone i le neurone j

3- Déclaration de tous les éléments de l'ensemble d'entrée (les trois paramètres topologiques)

4- Calcul de la fonction sigmoïde :

$$s(x) = \frac{1}{1 + e^{-x}}$$

$$ds(x) = s(x)(1 - s(x))$$

5- Calcul de l'erreur de ce neurone par soustraire les poids désirés des poids obtenu:

$$D_i = (S_i - Y_i)$$

6- Modification des poids du réseau par la règle :

$$W_{ij}(k) = W_{ij}(k - 1) - \text{pas} \times D_i \times O_j \times ds(I_i)$$

7- Calcul de l'erreur pour le neurone i de la couche cachée :

$$D_i = \sum_h D_h \cdot W_{hi}$$

8- Modification des poids suivant :

$$W_{ij}(k) = W_{ij}(k - 1) - \text{pas} \times D_i \times O_j \times ds(I_i)$$

9- calcul de l'erreur quadratique :

$$E = \sum_i (S_i - Y_i)^2$$

10- Incrémentation du numéro du neurone i.

11- Aller à l'étape 4 et répéter les calculs jusqu' à ce que E < seuil.

Figure 3.11: étapes d'apprentissage du réseau PMC

3.7.2 Méthode de classification

L'approche de classification adoptée, est un pseudo-arbre de classification. En effet le réseau SVNN, fait la séparation entre 3 classes bien distinctes, les masses en général, les calcifications et, les cas, sains.

Les 2 premières classes, sont séparées en deux sous classe distinctes, la sous classe maligne et la sous classe bénigne, comme le montre le schéma suivant:

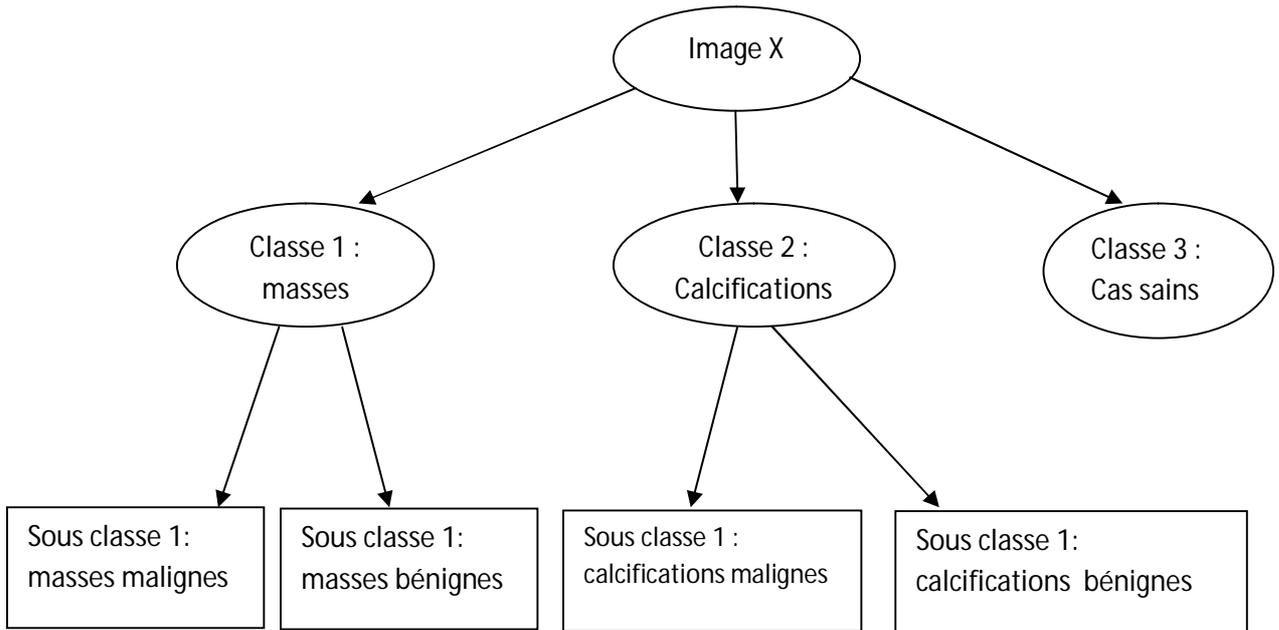


Figure 3.12 :l'approche de la classification et devisions des classes

3.8 Evaluation du résultat

Le résultat de la méthodologie adoptée, peut être évalué suivant diverses manières:

3.8.1 Courbe roc [43]

La courbe ROC (**Receiver Operating Characteristic Curve**), a été créé au cours de recherches, pour établir des signaux radios au milieu du bruit). On peut en générer selon cette évaluation plusieurs facteurs dont: la spécificité et la sensibilité.

- la sensibilité est : $TP / (TP + FN) = TP / P$ 3.19

- la spécificité est : $TN / (TN + FP) = TN / N$ 3.20

Avec;



- TP (true positive): les zones positives qui le sont vraiment.
- FP (false positive) : les zones positives, qui sont en fait négatives.
- TN (true négative): les zones négatives, qui le sont vraiment.
- FN (false négative): les zones négatives, qui sont en fait positives.
- P (positive): facteur représentant tous les positifs quel que soit l'état de leur prédiction, $P = TP + FN$.
- N (négative) : facteur représentant, tous les négatifs quel que soit l'état de leur prédiction. $N = TN + FP$.

La courbe ROC représente, l'évolution de la sensibilité (taux de vrais positifs) en fonction de la spécificité (taux de faux positifs) quand, on fait varier le seuil t . Au fur et à mesure que t augmente: la spécificité augmente, mais la sensibilité diminue.

C'est une courbe croissante entre le point (0,0) et le point (1, 1) et en principe au-dessus de la première bissectrice

3.8.2 matrice de confusion [35]

La matrice de confusion, dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente, le nombre d'occurrences d'une classe réelle (ou de référence). Les données utilisées, pour chacun de ces groupes doivent être différentes.

Les matrices d'erreurs (confusion) permettent, d'évaluer l'intensité de la liaison entre des données de référence et, le résultat d'une classification.

3.8.3 Distance de mahalanobis

La distance de Mahalanobis est une mesure de distance introduite par « **Prasanta Chandra Mahalanobis** [37]. Elle est basée sur la corrélation entre des variables par lesquelles différents modèles peuvent être identifiés et analysés C'est une manière utile de déterminer la similarité entre une série de données connues et inconnues. Elle diffère de la distance euclidienne par le fait qu'elle prend en compte la corrélation de la série de données. Ainsi, à la différence de la distance euclidienne où toutes les composantes des vecteurs sont traitées



de la même façon, la distance de Mahalanobis accorde un poids moins important aux composantes les plus bruitées (en supposant que chaque composante soit une variable aléatoire de type gaussien)

3.9 Conclusion

Ce chapitre a fait l'objet d'une analyse texturale, qui se base sur les méthodes statistiques, pour évaluer différents paramètres qui nous informent sur la distribution des niveaux de gris ainsi que l'aspect des pathologies mammaires. Une description morphologique, a également été prise en compte suivant la compacité, la surface et d'autres attributs, afin d'identifier le changement de la forme, dans le cas des images mammographiques. La discrimination de masses et de calcifications, est réalisée à partir des vecteurs d'attributs caractéristiques par une combinaison, entre le perceptron multicouche et l'algorithme génétiques.

Le chapitre suivant, présente quelques résultats engendrés par notre méthodologie, appliquée sur des images mammographiques réelles. Ces résultats sont suivis d'une discussion.



Chapitre 4

Application sur les images mammographiques et discussion des résultats

L'objectif de ce chapitre, est de présenter l'application développée, les traitements illustrés par des images mammographiques dans le but de l'extraction et de la classification des pathologies mammaires issues de la base de données MIAS (Mammographic Image Analysis Society) [45], suivant le perceptron multicouche amélioré par les processus génétiques. Ces images de type MLO (incidence latérale oblique) sont numérisées avec une résolution de 50 μm par pixel sur 8 bits. La résolution spatiale de chaque image est de 1024 x 1024 pixels. Les images étudiées, comprennent des images normales, des masses ainsi que des calcifications; les deux dernières catégories sont réparties suivant, des cas bénins et malins. La zone suspecte, est identifiée par des radiologues experts [45].

4.1. Environnement de travail

Notre système a été développé sous environnement Windows, à l'aide de l'outil de programmation matlab de la version 8.1 (R1013a).

4.1.1 Matériel utilisé

L'implémentation de notre système a été réalisée grâce à, un **PC** portable **Intel core I5-240M CPU @2.30GHz**, **CPU @2.30GHz** et, d'une capacité de mémoire vive «**4.04GB RAM** ».



4.1.2 Qu'est ce que Matlab?

Notre système a été développé sous environnement Windows, à l'aide de l'outil de programmation matlab (abréviation de " **Matrix Laboratory** ") de la version 8.1 [46], qui est un environnement informatique conçu, pour le calcul matriciel. L'élément de base est une matrice dont la dimension, n'a pas à être fixée. Matlab est un outil puissant, qui permet la résolution de nombreux problèmes en beaucoup moins, de temps. On peut se rendre compte de la puissance du logiciel, en lançant la commande *demo*. Les compléments d'information, peuvent être facilement obtenus, grâce à l'aide en ligne. Matlab s'élargit sur la librairie des fonctions mathématiques, l'environnement graphique, ainsi que sur une interface de développement.

Les utilisateurs de MATLAB, sont de domaines très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel, que pour la recherche.

Matlab fonctionne suivant deux modes; le mode interactif et, le mode de programmation.

Dans les deux cas, l'utilisateur peut définir ses propres fonctions et les utiliser.

- **Le mode interactif:** exécute des instructions intégrées, par l'utilisateur dans la fenêtre de commande de matlab (le workspace);
- **le mode de programmation:** exécute le programme ou script, qui sont des fichiers regroupant, une suite d'instructions (les fichiers .m).

4.1.3 Débuter en Matlab

Pour découvrir matlab d'une façon efficace, il faut utiliser son aide en ligne. On présente ci-dessous, quelques commandes intéressantes:

- **help** : « help » donne la liste des aides générales possibles;
- **helpwin** : ouvre une fenêtre et donne accès, à une aide détaillée;
- **help+ nom de commande** : donne l'utilisation de cette commande;
- **démo** : lance une démo générale de matlab;
- **Clear** : supprime la variable de l'espace de travail;
- **Clc** : efface l'écran son toucher les variables;
- **Exit** : permet de sortir de matlab;

- **Who** : donne le nom de variable présente dans l'espace de travail;
- **Look for+commande** : donne une Liste de commandes, qui ont un rapport avec le nom de commande.

S'il est parfaitement adapté à l'automatique et au traitement du Signal, sa facilité d'emploi avec des nombres complexes et ses possibilités d'affichages graphiques, en font un outil intéressant pour plusieurs types d'applications. De plus, des " toolboxes " (boîtes à outils) sont disponibles dans de nombreux domaines du traitement du signal, de la commande des systèmes, des réseaux de neurones, de la logique floue, des ondelettes, de la simulation, etc...

4.1.4 Présentation du 'gui'

Les interfaces graphiques (ou interfaces homme-machine) sont appelées GUI (Graphical User Interface) sous MATLAB. Elles permettent à l'utilisateur, d'interagir avec un programme informatique, grâce à différents objets graphiques (boutons, menus, cases à cocher, ...). Ces objets sont généralement, actionnés à l'aide de la souris ou du clavier. Leur efficacité et leur ergonomie, sont essentielles dans l'acceptation et l'utilisation de ces outils par les utilisateurs. Le GUIDE (cf. Figure 4.1), est un constructeur d'interface graphique qui regroupe tous les outils dont, le programmeur a besoin pour créer une interface graphique de façon intuitive. Il s'ouvre, en tapant **guide** dans la commande « Window » de MATLAB.

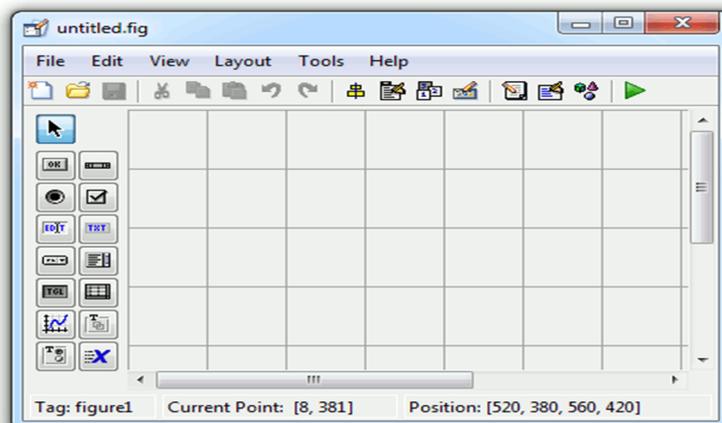


Figure 4.1: fenêtre principale du guide

Un double-clic sur un objet permet de faire apparaître, le « Property Inspector », où les propriétés des objets sont facilement éditables. Leurs modifications et la visualisation de ces modifications, sont immédiates (cf. Figure 4.2).

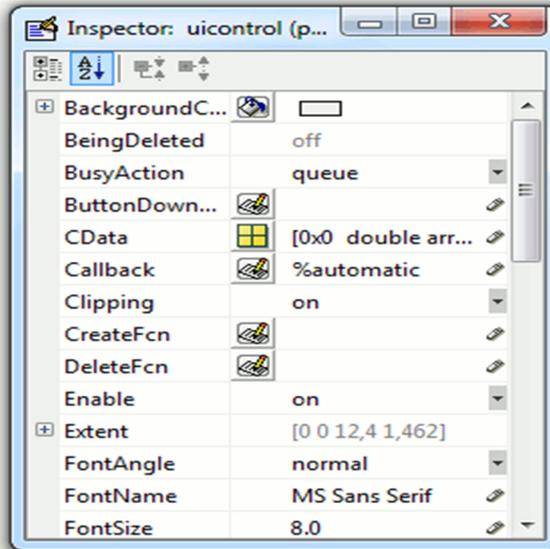


Figure 4.2 : Property Inspector

Certain « Handlers » sont définis par défaut :

- **gca** : axes courants de tracés (get current axes);
- **gcbf** : la figure cliquée (get callback figure);
- **gcbo** : l'objet qui appelle (celui sur lequel on a cliqué);
- **gcf** : handler de la figure courante (get current figure).
-

Certaines instructions nécessitent un handler;

- **get** et **set** : **permet** respectivement de lire et changer les valeurs des propriétés;
- **findobj** : permet de retrouver le handler d'un objet;
- **propedit** : permet d'éditer les propriétés d'un objet ;
- **delete** : permet d'effacer un objet.



4.2 Présentation du système développé

Le système (cf. Figure 4.3) que nous avons développé a pour objectif, de classifier les masses et les calcifications mammographiques moyennant les SVNN, qui utilisent le perceptron multicouche dont, les paramètres neuronaux sont ajustés par les algorithmes génétiques, pour réduire les confusions. L'identification de ces zones s'est faite, par les étapes suivantes :

1. en raison de la variabilité des tissus mammaires, d'une patiente à l'autre, l'extraction de la zone d'intérêt se fait d'une manière manuelle, moyennant l'annotation de l'expert;
2. les applications des algorithmes de segmentation de la zone d'intérêt se fait grâce à la coopération de deux techniques;
 - coopération des contours actifs et de la segmentation par analyse d'histogramme pour l'extraction des masses;
 - coopération des contours actifs et de la segmentation par analyse d'histogramme pour l'extraction des calcifications et post traitement détectées par le chapeau haut de forme.
3. étiquetage simple, de la région segmentée, pour calculer le nombre des régions ;
4. extraction des caractéristique texturaux et morphologiques, pour chaque région étiquetée;
5. classification, effectuée par le classifieur **SVNN** (cascadage entre deux classifieurs **(RN)** et **(AG)**) qui permet de discriminer suivant la description texturale et morphologique; les tissus normaux et les tissus anormaux.

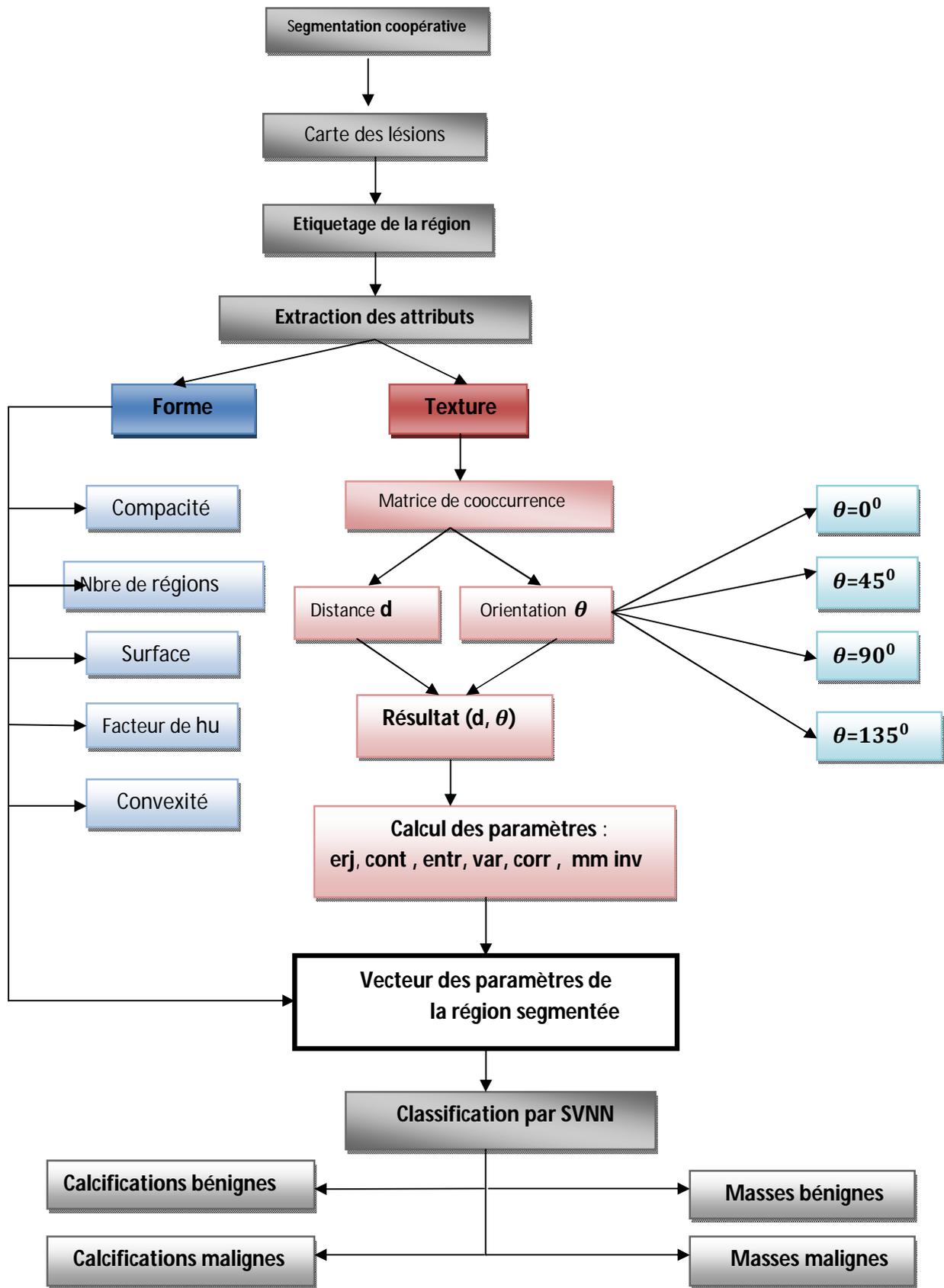


Figure 4.3 : système d'aide à la décision mammographique

4.2.1 Images utilisées

Les différentes pathologies mammaires issues de la base de données MIAS (Mammographic Image Analysis Society) [45], qui font l'objet de notre étude, illustrent principalement des calcifications bénignes et malignes, des masses bénignes et malignes et, des cas sains. Nous nous sommes intéressés à l'analyse d'une zone annotée, par des radiologues experts.

La base (MIAS) est une base de données de mammographies numériques. Nous avons considéré 152 images dans notre étude. La figure suivante illustre quelques exemples de la base MIAS.

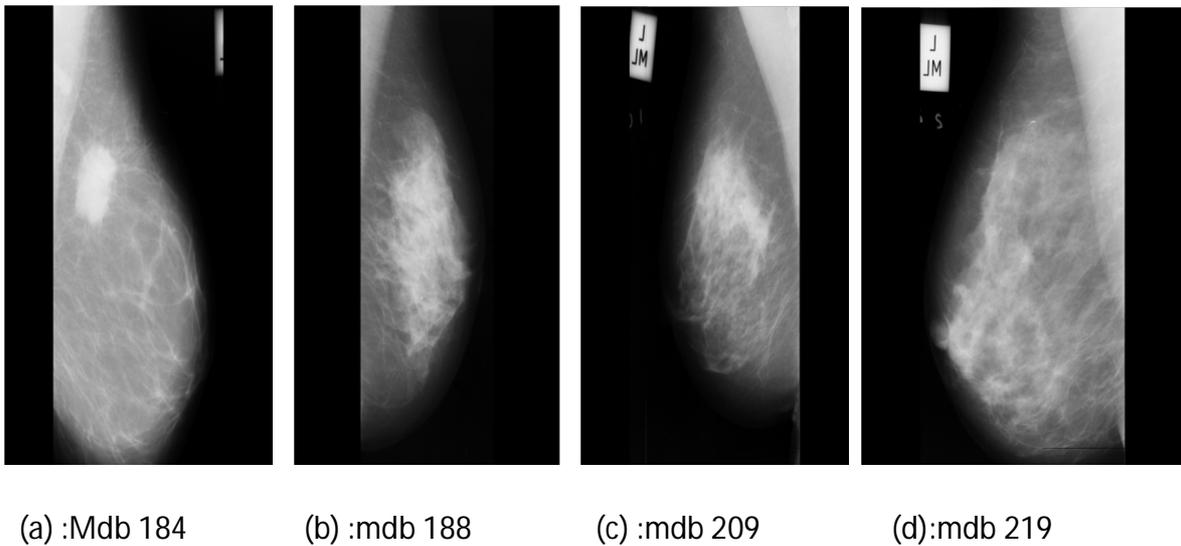


Figure 4.4: image mammographiques de la base MIAS

(a): masse maligne, (b) :masse bénigne , (c) :calcifications malignes, (d) :calcifications bénignes

4.2.2 Interface du système

La figure 4.5, présente l'interface de notre application, que nous appelons (SVNN mammo). Elle permet plusieurs traitements suivant les divers menus:

- **fichier** : ce module permet d'ouvrir les image et les fichiers;
- **segmentation** : **permet de** réaliser les traitements ci-dessous :

Région d'intéret : introduire la position et le rayon de la tumeur,

option de segmentation : sélectionne la méthode de segmentation et exécute

L'algorithme de segmentation,

Figure : permet d'illustrer les figures de chaque segmentation.

➤ **Caractérisation** : effectue la caractérisation des régions d'intérêt suivant les

Paramètres de texture et de forme

➤ **Classification** : ce module permet de classer le type de tumeur (masse, calcifications cas sains) et d'introduire le taux de classification ainsi que la matrice de confusion.

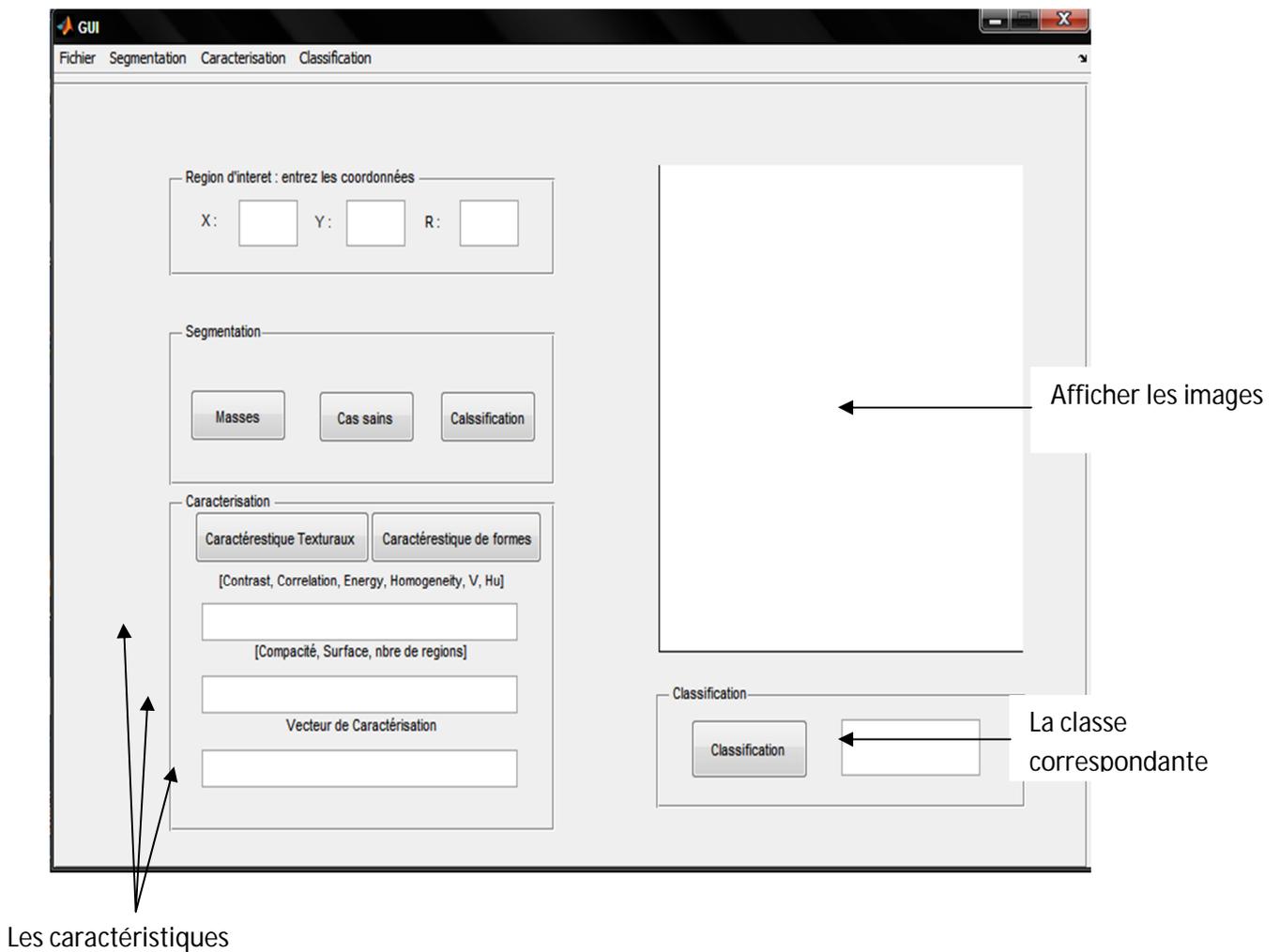


Figure 4.5 : interface de l'application



4.3 Résultats et interprétation

L'application (SVNNmammo) parcourt l'analyse texturale et topologique des masses, des calcifications et des cas sains, en vue d'une identification de ces données, pour l'aide à l'interprétation.

4.3.1 Détection des masses

La détection des masses s'est faite, par une technique de segmentation coopérative moyennant, les contours actifs pour l'obtention de la forme de la masse et, la segmentation par analyse d'histogramme, pour l'isolation de la région.

Le choix du nombre d'itérations a été pris en compte judicieusement, puisque lors de nos tests, nous avons remarqué que la taille de la masse (telle que expertisée par un médecin spécialiste), influe sur le nombre d'itérations considéré. Le tableau ci-dessous illustre les nombre d'itérations et, les seuils pour quelques images de la base.

	Mdb 184	Mdb 270	Mdb271	Mdb178	Mdb171	Mdb125	Mdb124
N. itérations	90	100	130	150	100	100	110
seuil	155	155	155	155	155	155	155
Rayon	114	72	68	170	62	60	33

Tableau 4. 1 : paramètres de segmentation de masses malignes

	Mdb 005	Mdb 080	Mdb091	Mdb099	Mdb144	Mdb199	Mdb290
N. itérations	250	15	100	100	100	50	50
Seuil	155	155	155	155	155	155	155
Rayon	30	20	20	23	29	31	45

Tableau 4. 2 : paramètres de segmentation de masses bénignes

Un exemple de résultat de segmentation de l'image mdb184, est illustré par les figures suivantes:



Figure 4.6 : détection d'une masse maligne par les contours actifs

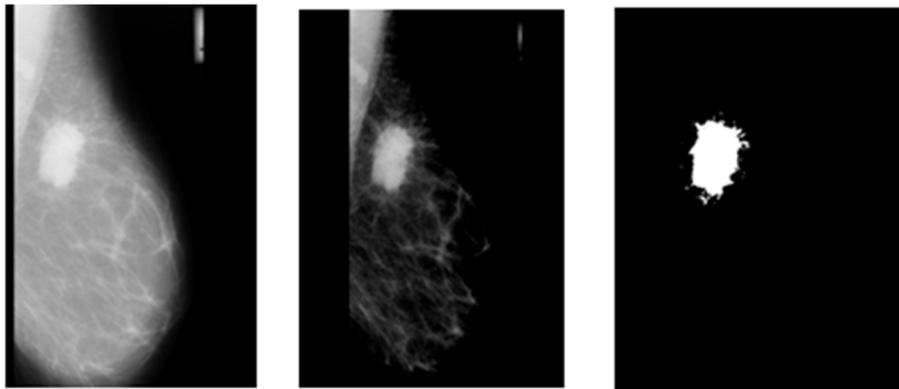


Figure 4.7 : détection d'une masse maligne par analyse d'histogramme



Figure 4.8 : détection d'une masse maligne par coopération

Nous remarquons que ces traitements, ont en effet produit un nombre d'itérations selon la taille de la masse considérée. Ainsi, dans le cas de l'image mdb 184, la surface est moins élevée, que pour le cas de l'image mdb 214, qui présente une masse assez importante. Le



nombre d'itérations, pour l'obtention du contour, est donc plus réduit pour l'image mdb 184.

4.3.2 Détection de calcifications

La détection de calcifications se fait suivant, une coopération entre les modèles déformables et la segmentation para analyse d'histogramme. Un post traitement est effectué sur le résultat de la coopération, moyennant une opération de la morphologie mathématique, produite par le chapeau haut forme, isoler les calcifications des autres structures. L'élément structurant utilisé pour le chapeau haut de forme correspond à un disque avec une taille de 5X5, car c'est la taille moyenne, d'un élément de base d'une calcification.

Le tableau ci-dessous, illustre les nombre d'itérations, les seuils ainsi que la taille de l'élément structurant adopté pour tous les cas considérés.

	Mdb214	Mdb218	Mdb222	Mdb227	Mdb236	Mdb248	Mdb252
N. itérations	50	60	220	150	50	50	50
seuil	155	155	155	155	155	155	155
C.H.F	5*5	5*5	5*5	5*5	5*5	5*5	5*5
Rayon	87	13	45	40	38	64	64

Tableau 4.3 : paramètres de segmentation de calcifications malignes

	Mdb214	Mdb218	Mdb222	Mdb227	Mdb236	Mdb248	Mdb252
N. itérations	50	150	150	50	150	160	50
seuil	155	155	155	155	155	155	155
C.H.F	5*5	5*5	5*5	5*5	5*5	5*5	5*5
Rayon	15	25	30	40	44	25	20

Tableau 4.4 : paramètres de segmentation de calcifications bénignes

Dans le cas des calcifications, le nombre d'itérations pour la détection du contour actif, augmente en fonction de la taille zone d'intérêt.

Un exemple de résultat de segmentation de l'image mdb209 qui représente le cas de calcifications malignes, est illustré par les figures suivantes:

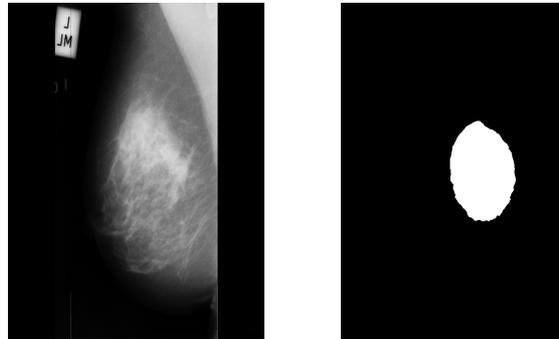


Figure 4.9 : détection de la forme des calcifications par les contours actifs (mdb 209, nombre d'itérations=1500)

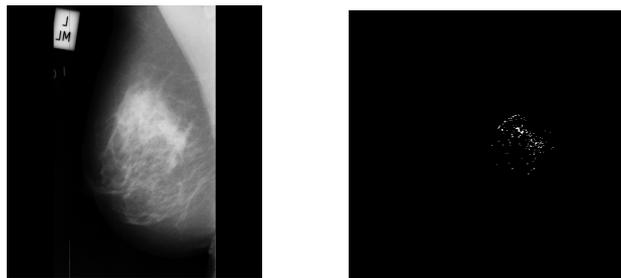


Figure 4.10 : détection des calcifications par analyse d'histogramme (seuil= 150)

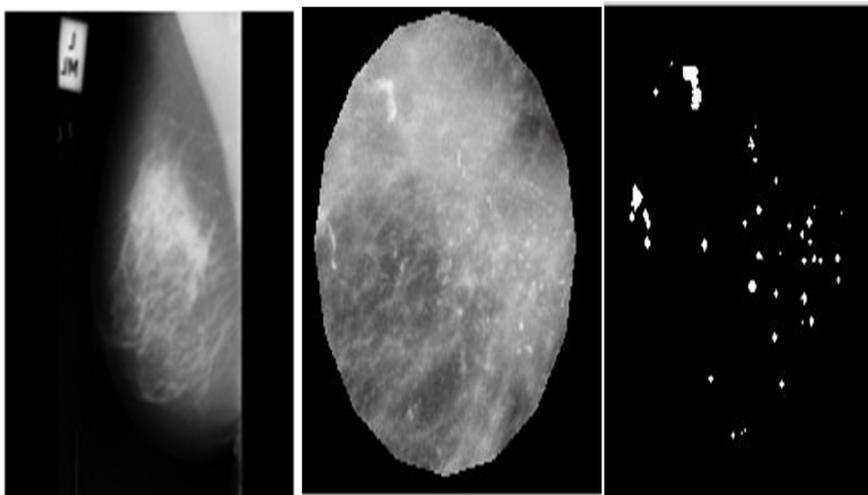


Figure 4.12 : détection de calcification par coopération et isolation par chapeau haut de forme

4.3.3 Extraction des caractéristiques

La base mammographique MIAS répartie sur cinq classes, suivant des masses malignes, et bénignes, des calcifications malignes et bénignes, ainsi que, des cas sains a été analysée en explorant six paramètres texturaux (contraste, corrélation, énergie, homogénéité, variance, entropie) et cinq attributs morphologiques (compacité, nombre de régions, surface, moment de hu). L'objectif, est de réaliser une base de connaissances via, cette description.

Les tableaux suivants, représentent l'évaluation des paramètres utilisés pour la caractérisation, de chaque classe.

	Mdb 184	Mdb 270	Mdb271	Mdb178	Mdb171	Mdb125	Mdb124
Contraste	0.010623	0.0078774	0.006224	0.0072703	0.0072894	0.0065009	0.0059033
Corrélation	0.99096	0.98997	0.99218	0.99043	0.99023	0.99194	0.99181
Energie	0.97167	0.96262	0.97389	0.96759	0.96524	0.96486	0.96515
Homogénéité	0.99955	0.99913	0.99949	0.99917	0.99913	0.99921	0.99922
Variance	2.2199e+21	2.1787e+21	2.2302e+21	2.2014e+21	2.1904e+21	2.1889e+21	2.1903e+21
Entropie	0.18428	0.25196	0.17899	0.22335	0.22495	0.24268	0.24583
Compacité	1.1042	1.0953	1.108	1.2283	1.2002	1.1709	1.1874
Nbre de régions	1	1	1	1	1	1	1
Surface	14857	19669	13671	17001	18275	18463	18304
M. Hu	0.025725	0.035616	0.029114	0.033876	0.035723	0.033799	0.034785

Tableau 4.5: paramètres des images de la classe des masses malignes

	Mdb 005	Mdb 080	Mdb091	Mdb099	Mdb144	Mdb199	Mdb290
Contraste	0.0015388	0.00084005	0.001766	0.0015274	0.0017431	0.0019417	0.0021822
Corrélation	0.96726	0.96346	0.96561	0.96781	0.97266	0.96397	0.96556
Energie	0.99802	0.99855	0.99788	0.99804	0.99738	0.99832	0.99816
Homogénéité	0.99992	0.99992	0.99994	0.99995	0.99993	0.99993	0.99994
Variance	2.3426e+21	2.3451e+21	2.3419e+21	2.3427e+21	2.3396e+21	2.344e+21	2.3433e+21
Entropie	0.015118	0.012509	0.015748	0.014972	0.020252	0.013408	0.014297
Compacité	1.0789	1.0166	1.0285	1.0321	1.0731	1.0197	1.0279
Nbre de régions	1	1	1	1	1	1	1
Surface	1004	733	1077	995	1338	851	934
M. Hu	0.036161	0.038978	0.032645	0.033161	0.03444	0.029266	0.028462

Tableau 4.6: paramètres des images de la classe des masses bénignes

	Mdb 209	Mdb211	Mdb213	Mdb239	Mdb241	Mdb249a	Mdb249b
Contraste	0.008515	0.0018691	0.0029555	0.0034729	0.0017049	0.003975	0.0034003
Corrélation	0.5316	0.7716	0.86539	0.87592	0.89733	0.60946	0.60161
Energie	0.9992	0.99961	0.99882	0.99886	0.9992	0.99948	0.99967
Homogénéité	0.9997	0.99993	0.99987	0.99988	0.99993	0.99986	0.99992
Variance	2.3485e+21	2.3501e+21	2.3463e+21	2.3466e+21	2.3482e+21	2.3494e+21	2.3504e+21
Entropie	0.003788	0.0030913	0.0090516	0.008655	0.0063705	0.0035127	0.0022543
Compacité	16.828	3.0279	5.1963	2.9845	1.7617	13.511	6.5304
Nbre de régions	50	11	12	17	12	26	21
Surface	258	170	549	536	379	192	123
M. Hu	1.1032	0.11007	0.32235	0.098833	0.045654	63.121	6.1742

Tableau 4.7: paramètres des images de la classe des calcifications malignes

	Mdb214	Mdb218	Mdb222	Mdb227	Mdb236	Mdb248	Mdb252
Contraste	0.0004087	0.0015503	0.001892	0.0012601	0.0014224	0.0020849	0.0029211
Corrélation	0.040353	0.36808	0.37117	0.38316	0.44401	0.67633	0.41015
Energie	0.99998	0.99991	0.99988	0.99992	0.99991	0.99981	0.99982
Homogénéité	0.99999	0.99997	0.99996	0.99997	0.99997	0.99996	0.99994
Variance	2.3518e+21	2.3514e+21	2.3515e+21	2.3516e+21	2.3515e+21	2.3511e+21	2.3511e+21
Entropie	0.0001513	0.0005021	0.0007468	0.00045134	0.0005263	0.0011135	0.0008693
Compacité	0.11595	1.4418	2.7467	2.8257	1.539	3.7577	4.4531
Nbre de régions	6	11	20	7	11	6	14
Surface	8	30	44	26	31	74	58
M. Hu	-2382.9	-744.55	-6223.4	0.25029	6.9386	0.080381	1.8438

Tableau 4. 8: paramètres des images de la classe des calcifications bénignes

	Mdb003	Mdb009	Mdb011	Mdb014	Mdb039	Mdb68	Mdb070
Contraste	0.033008	0.030861	0.030242	0.032142	0.020499	0.014472	0.016501
Corrélation	0.99536	0.99356	0.99361	0.99559	0.99119	0.99301	0.99154
Energie	0.8008	0.7216	0.80021	0.73253	0.92862	0.87676	0.92195
Homogénéité	0.99652	0.99395	0.99572	0.99642	0.9986	0.99833	0.99903
Variance	1.4837e+21	1.1764e+21	1.4667e+21	1.2148e+21	2.0241e+21	1.7926e+21	1.9919e+21
Entropie	1.1048	1.4974	1.0072	1.4529	0.4172	0.66351	0.4454
Compacité	1.9578	1.4745	1.8166	1.5868	1.7392	2.1502	2.4265
Nbre de régions	1	1	1	1	1	1	1
Surface	1.1285e+05	1.6486e+05	1.1533e+05	1.5783e+05	38140	68081	42144
M. Hu	0.042455	0.043422	0.036961	0.031666	0.044871	0.016501	0.0372

Tableau 4.9: paramètres des images de la classe des cas sains

L'extraction de paramètres topologiques, nous permet de constituer des vecteurs caractéristiques de chacune des formes étudiées, en vue d'une classification connexionniste, pour une prise de décision. Nous remarquons que dans le cas d'images saines, les paramètres descriptifs, sont du même ordre, alors que pour les cas pathologiques, ils varient suivant le type de lésion.

Nous montrons à cet effet, l'influence de la surface et du contraste, sur les masses et les calcifications, suivant les graphes des figures 4.13, 4.14 ; 4.15 et 4.16.

a. Influence de la surface sur la masse

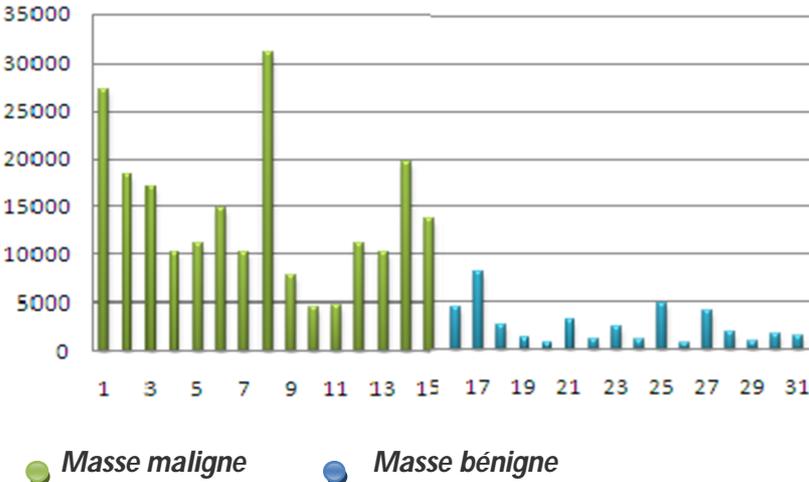


Figure 4.13: influence de la surface sur les masses

b. influence du contraste sur les masses

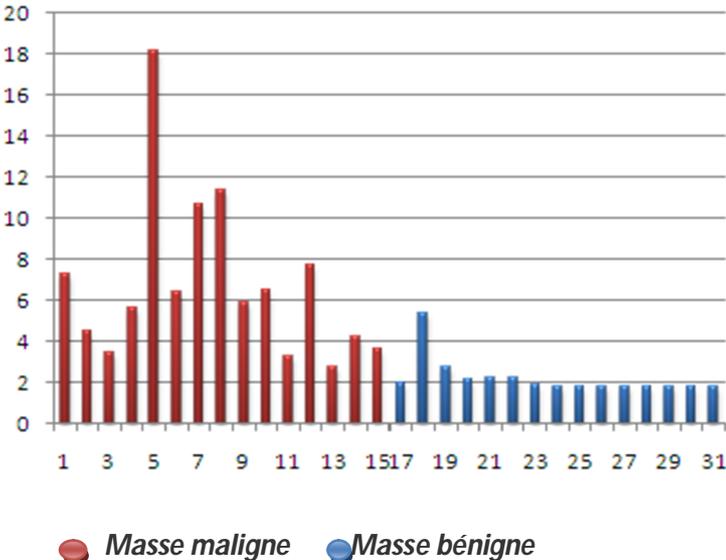


Figure 4.14: influence du contraste sur les masses



C. Influence de la surface sur les calcifications

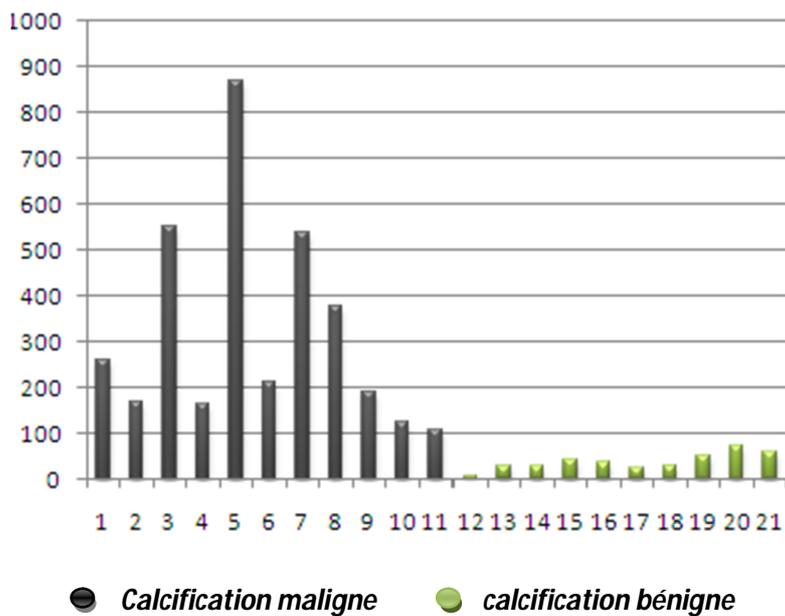


Figure 4.15 : influence de la surface sur les calcifications

d. Influence de la compacité sur les calcifications

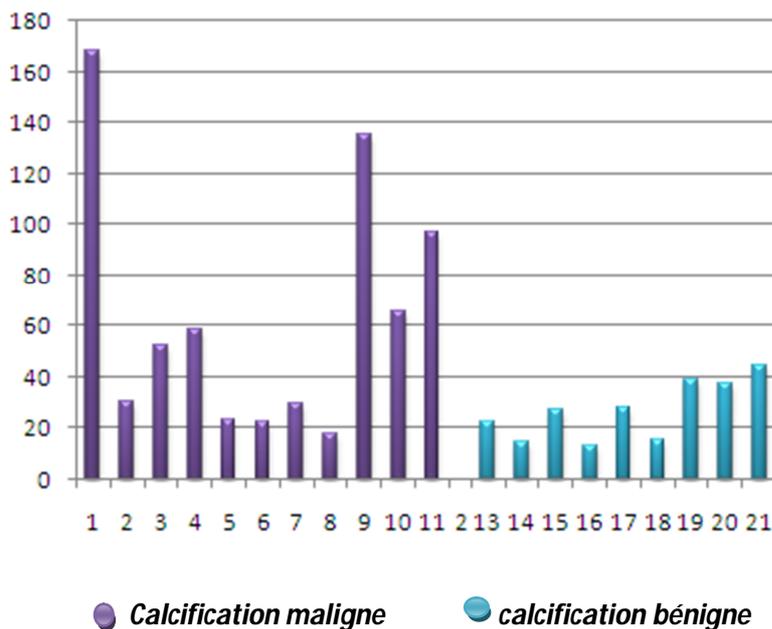


Figure 4.16: influence de la compacité sur les calcifications

Nous remarquons que lorsque les calcifications sont malignes, la compacité augmente ainsi que leur taille. Le contraste, est aussi un paramètre important, lorsqu'il s'agit d'une masse maligne.



4.3.4 Classification

Il n'existe pas de forme spécifique représentative des réseaux de neurones multicouches en dehors du fait qu'ils possèdent plusieurs couches internes. En particulier, le nombre de couches et le nombre de cellules par couche cachée, dépendent principalement du problème étudié et des résultats de l'expérimentation sur différentes combinaisons des valeurs de ces paramètres. Nous utilisons le modèle du perceptron multicouche en cascading avec les algorithmes génétiques, pour classifier les différentes pathologies mammaires étudiées.

L'apprentissage est l'étape essentielle dans les systèmes de classification. Les résultats corrects (c'est-à-dire les valeurs que l'on désire que le réseau obtienne en sortie) sont fournis au réseau, si bien que celui-ci peut ajuster ses poids de connexions pour les obtenir. Après l'apprentissage, le réseau est testé en lui donnant seulement les valeurs d'entrée mais pas les sorties désirées, et en regardant si le résultat obtenu est proche du résultat désiré. Nous considérons, la couche d'entrée qui correspond aux paramètres extraits, une couche cachée et une couche de sortie pour produire les cinq classes étudiées.

Les paramètres du classifieur hybride utilisés sont :

C=70; le paramètre de régularisation ;

Max _ indiv =1500; la population initiale,

Pression = 5 ; la pression sélective [1-8],

Score = 1 ; mettre (1=oui, 0=non),

max_resol =12; le nombre de résolutions de l'algorithme génétique,

iter_max = 10 ; le nombre de générations par résolution.

L'apprentissage a été effectué avec les deux tiers de la base considérée (une centaine d'images). Le graphe ci-dessous représente, le déroulement de l'apprentissage du réseau de neurones du type SVNN que nous avons utilisé. « Les meilleurs individus » représentés en rouge, sont initialisés à 1500 individus et, selon les besoins du réseau de neurones, ce nombre peut changer. Comme nous avons considéré 10 neurones (après des tests empiriques) dans la couche cachée, ce nombre diminue rapidement, juste après les

premières résolutions. Même si à chaque résolution, ils sont initialisés à 1500, ce chiffre s'affaiblit; c'est tout le principe de l'apprentissage.

La fonction fitness moyenne, est représentée en bleu, elle correspond à la « santé » des individus; le but est d'avoir les meilleurs individus possibles pour notre réseau.

Le nombre de générations, représente le nombre de fois que les étapes « mutation et croisement » sont faites sur les individus; 11 dans notre cas, on se retrouve avec un nombre $N+1$ de résolution, qui est égale donc à 12. Le déroulement de l'apprentissage, s'effectue avec 120 itérations.

Sans oublier que les individus ici, représentent les poids, et les biais du réseau.

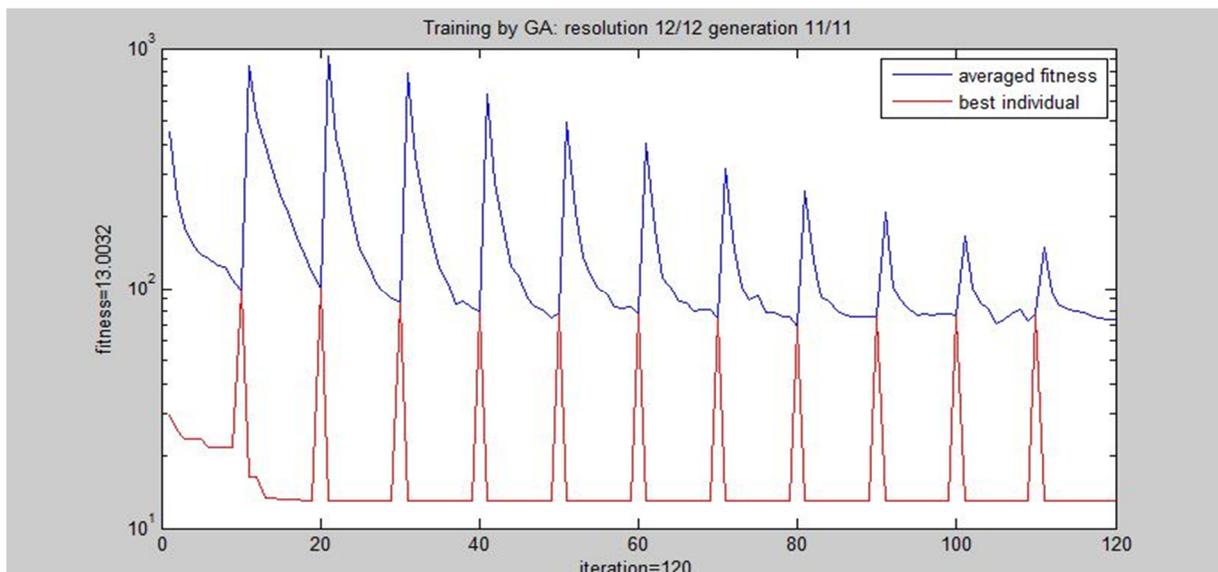


Figure 4.17: entraînement du réseau de neurones par les algorithmes génétiques

4.3.5 Evaluation des résultats

Dans cette phase nous faisons le test pour chaque classe, nous obtenons des résultats efficaces pour notre classifieur hybride. Nous avons alors, évalué le taux de reconnaissance, suivant la matrice de confusion montrée dans la figure 4.18.



	MM	MB	CM	CB	CS	ALL
Class MM	95%	0%	0%	0%	0%	95%
Class MB	0%	100%	0%	0%	0%	100%
Class CM	0%	0%	95%	5%	0%	100%
Class CB	0%	0%	5%	95%	0%	100%
Class CS	5%	0%	0%	0%	100%	105%
ALL	100%	100%	100%	100%	100%	97%

Figure 4.18 : matrice de confusion pour la catégorie des tests

La matrice de confusion est une manière de déterminer, l'efficacité du classifieur utilisé. Pour cette figure, nous avons utilisé la base des images tests (un tiers de la base), des images dont nous connaissons bien la nature mais, que, le SVNN ignore. La diagonale représente ainsi, les images reconnues dans leurs classes respectives, elle s'approche du résultat optimal, ce qui montre l'efficacité de notre réseau. Les autres cases, représentent les fausses identifications, comme on peut le constater, le taux de fausses classifications, est très bas.

Les figures ci-dessous illustrent des exemples de classification à partir de l'application 'SVNNmammo'. Le système développe engendre, un taux de classification de 97%.

➤ Cas de masses malignes

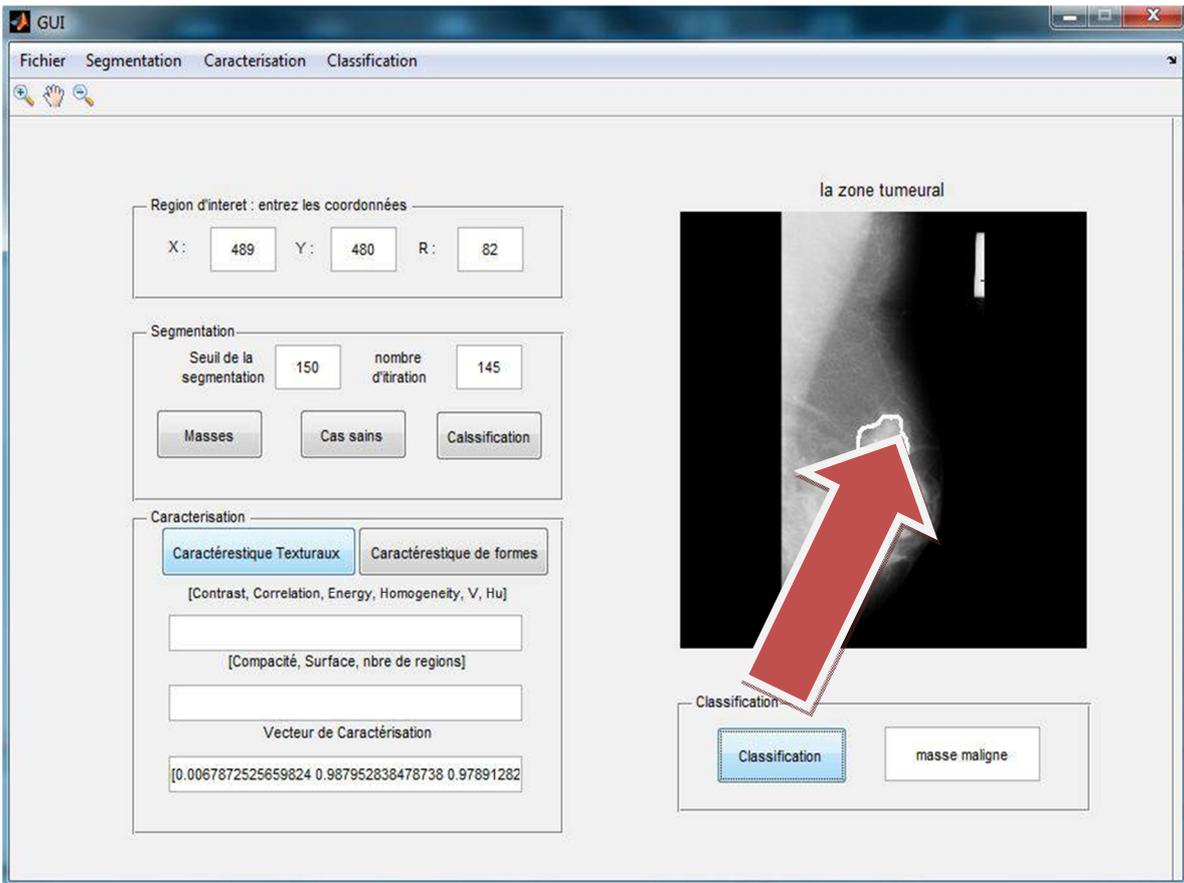


Figure 4.19 : classification d'une masse maligne (mdb274)

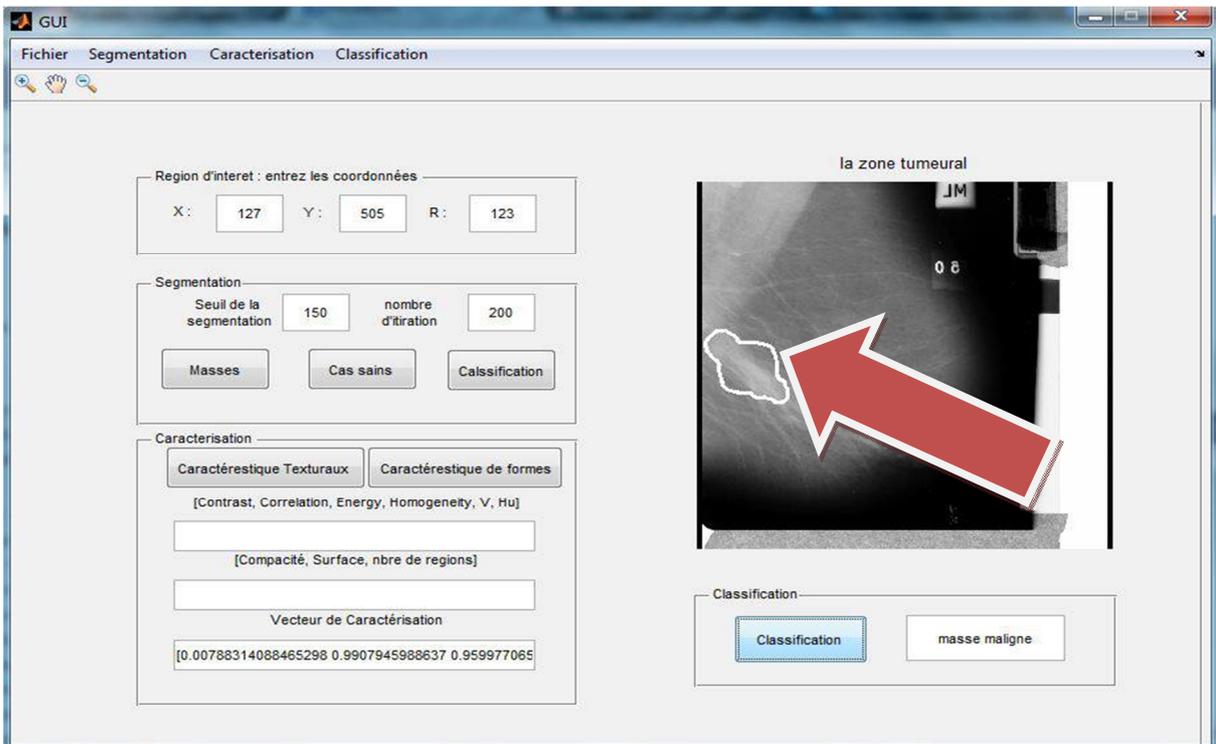


Figure 4.20 : classification d'une masse maligne (mdb170)

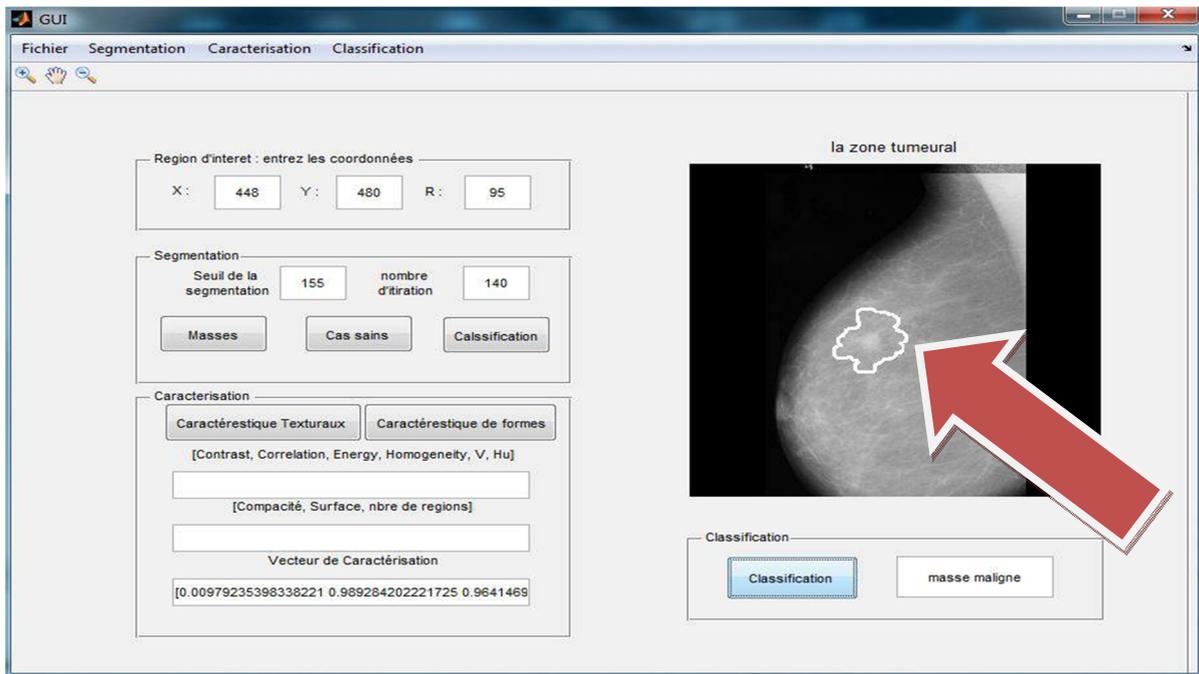


Figure 4.21 : classification d'une masse maligne (mdb155)

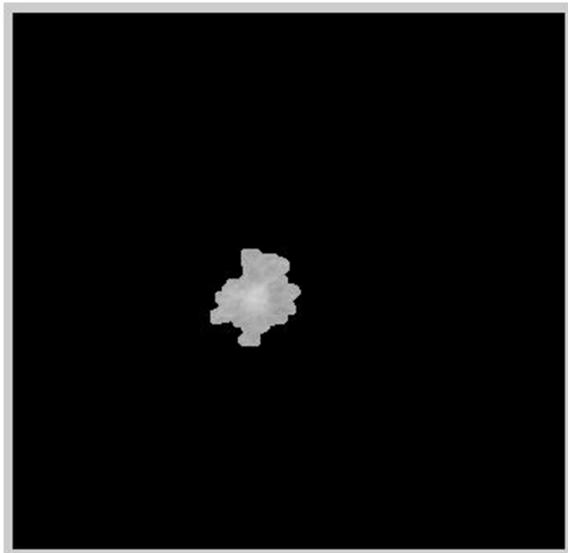
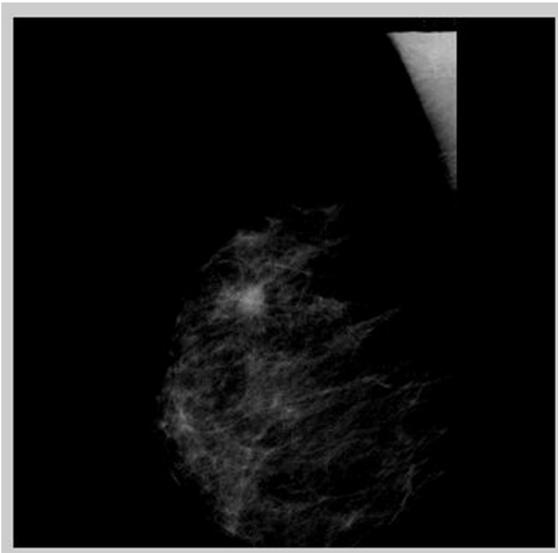


Figure 4.21.a : résultat de l'analyse de l'histogramme Figure 4.21.b : résultat du contour actif

Seuil utilisé= 150, nombre d'itérations=140

➤ Cas de masses bénignes

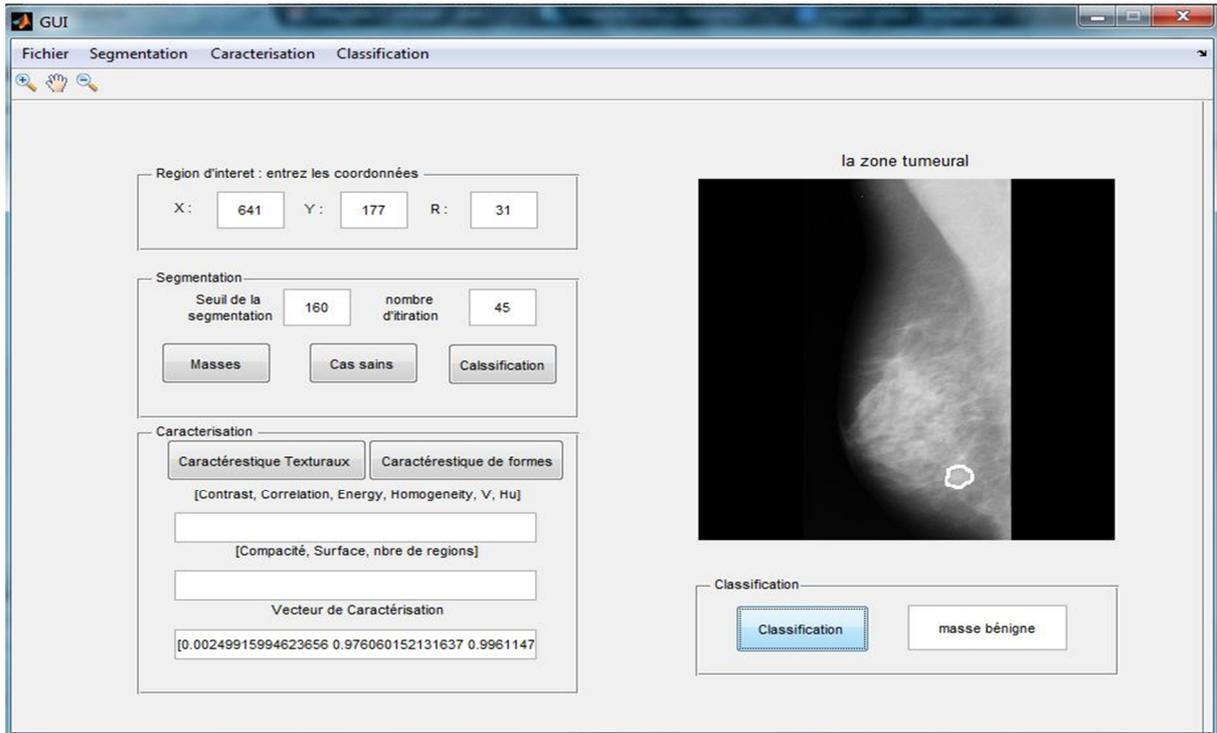


Figure 4.22 : classification d'une masse bénigne (mdb097)

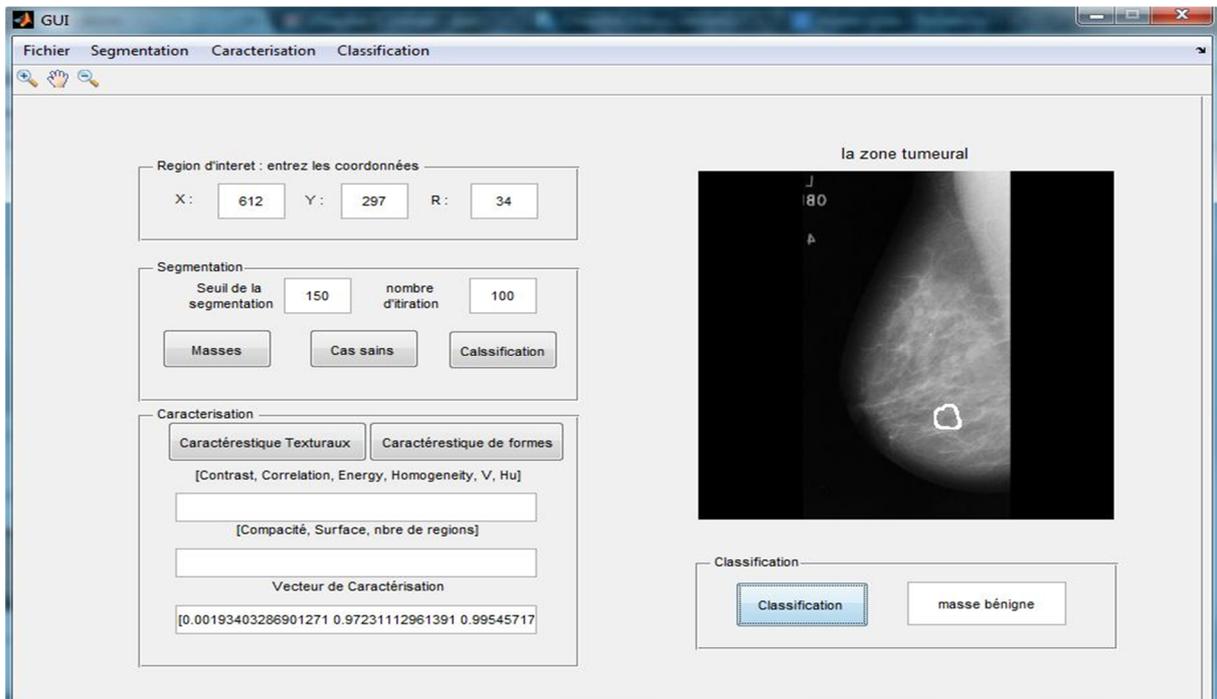


Figure 4.23 : classification d'une masse bénigne (mdb199)

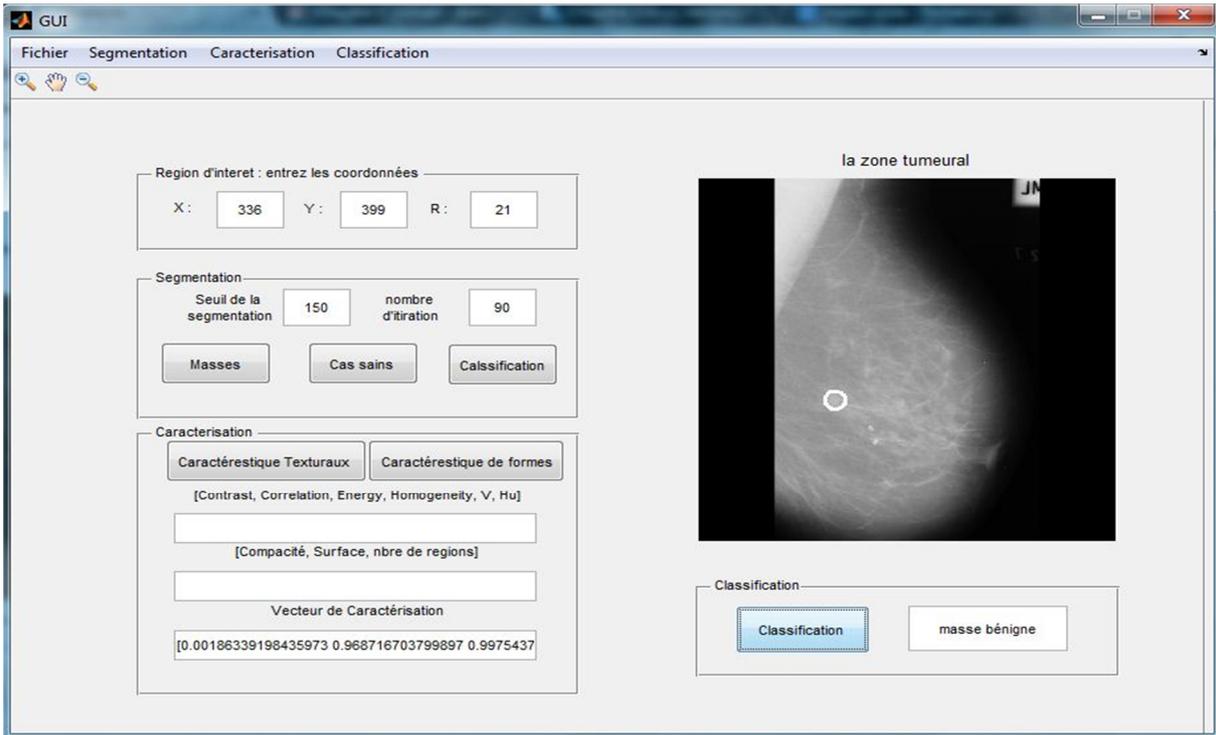


Figure 4.24 : classification d'une masse bénigne (mdb204)

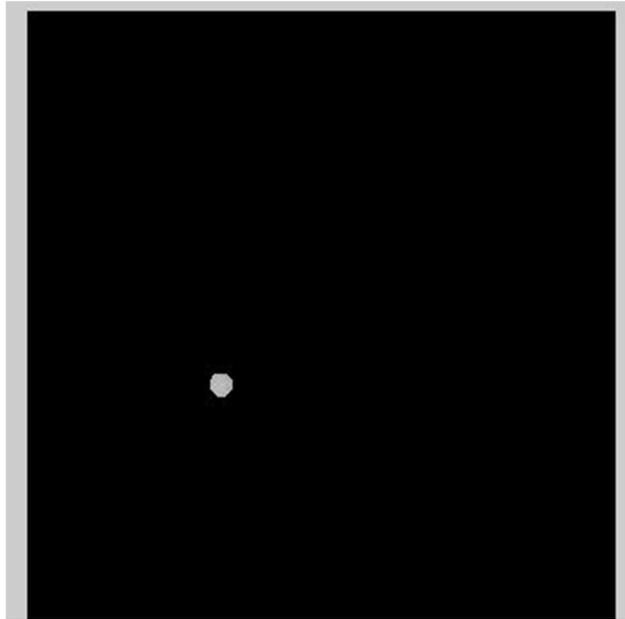


Figure 4.24.a : résultat de l'analyse de l'histogramme Figure 4.24.b : résultat du contour actif

Seuil utilisé= 150, nombre d'itérations=90

➤ Cas de calcifications bénignes

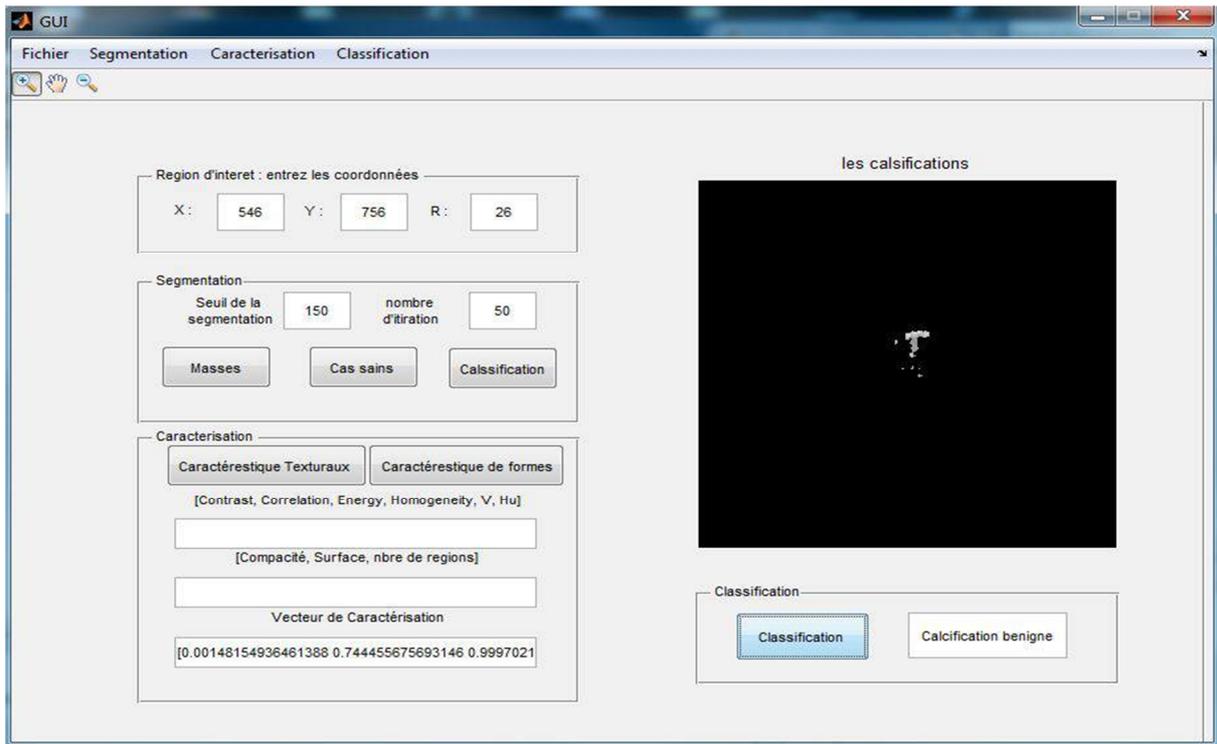


Figure 4.25: classification de calcifications bénignes (mdb 218)

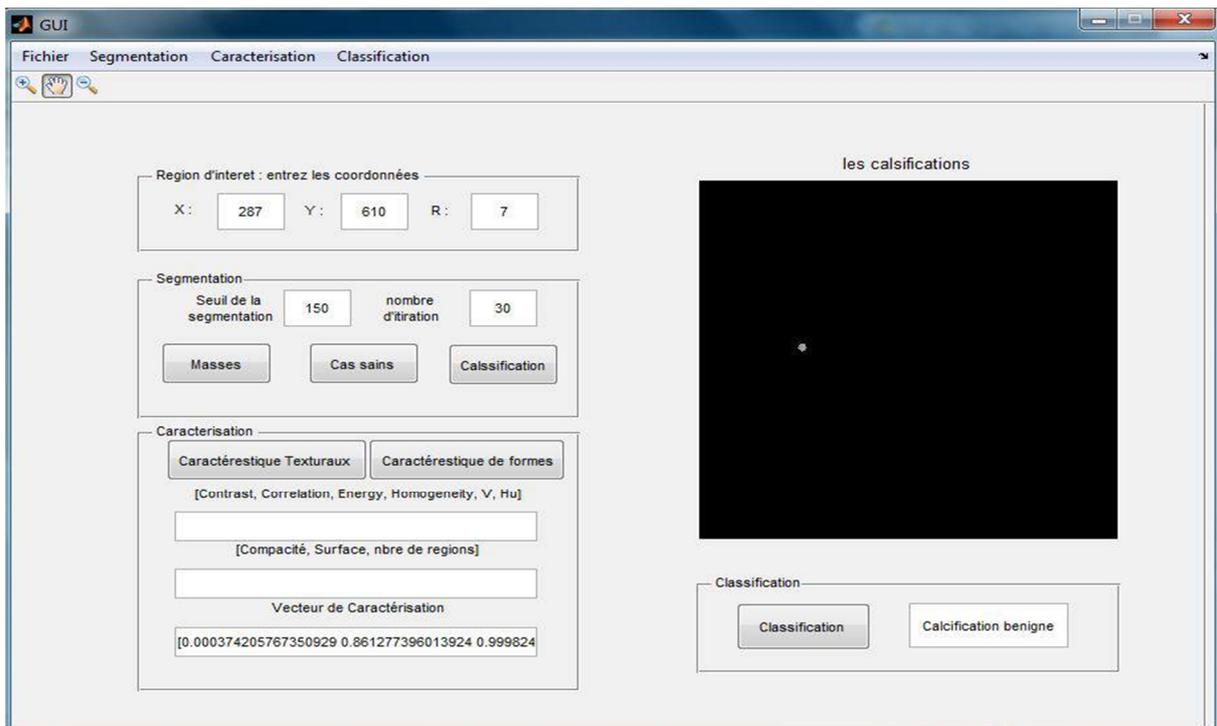


Figure 4.26: classification de calcifications bénignes (mdb219)

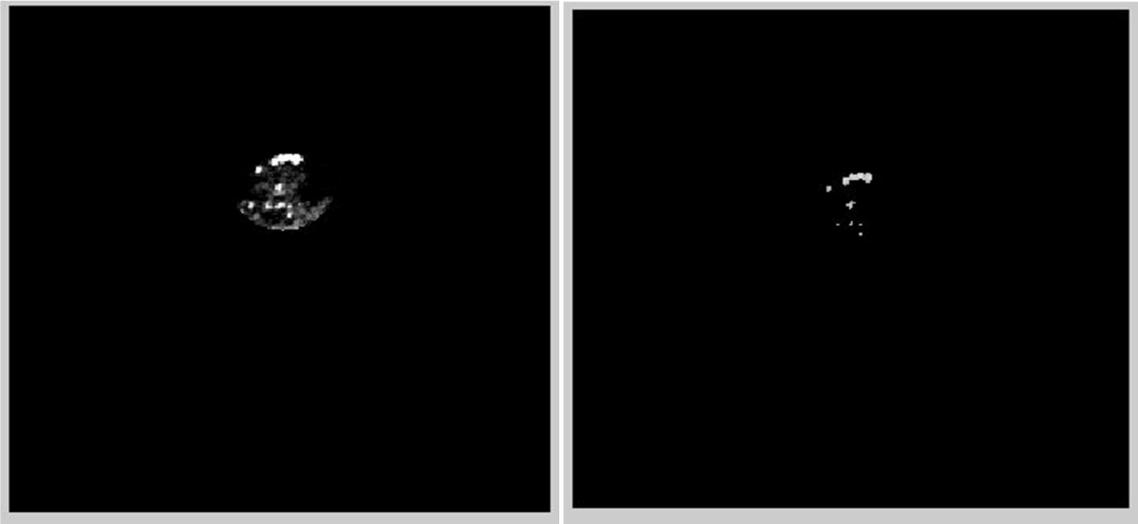


Figure 4.26.a : résultat de l'analyse de l'histogramme Figure 4.26.b : résultat du contour actif

Seuil utilisé= 150, nombre d'itérations=50

Ce choix du nombre d'itérations, est très petit par rapport, à celui utilisé par les calcifications malignes, car ici, c'est une petite zone, et peu de régions à extraire.

➤ **Cas de calcifications malignes**

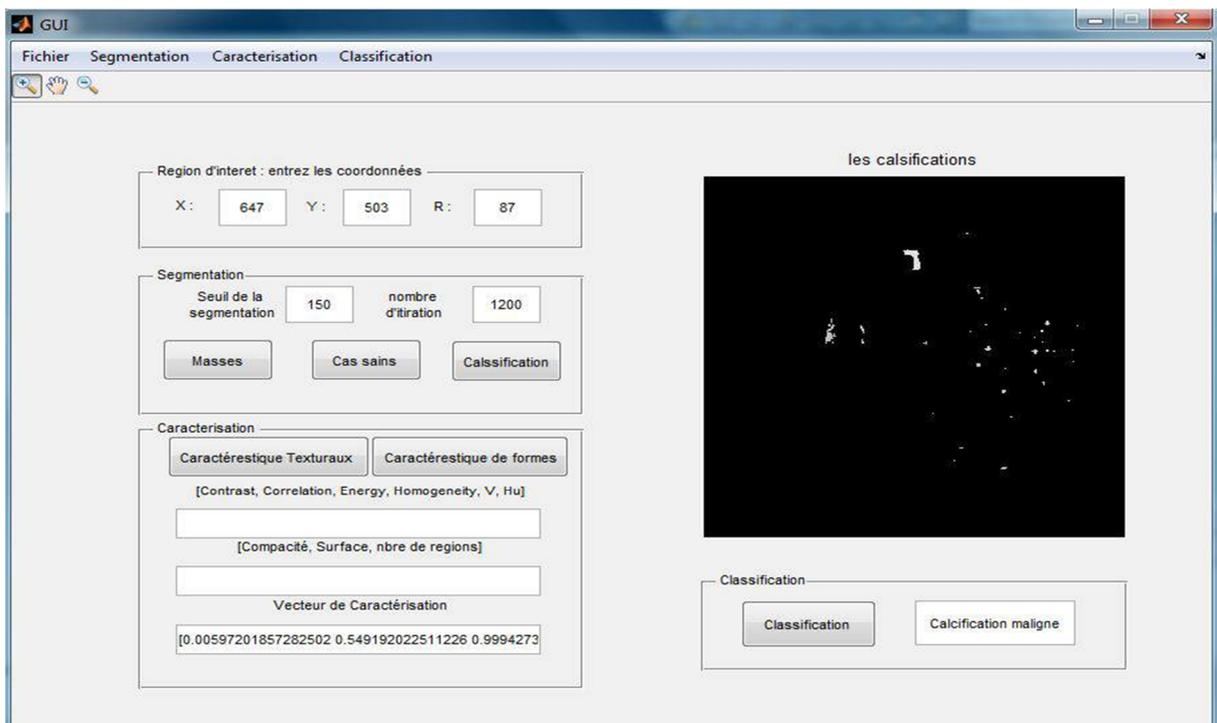


Figure 4.27: classification de calcifications malignes (mdb226)

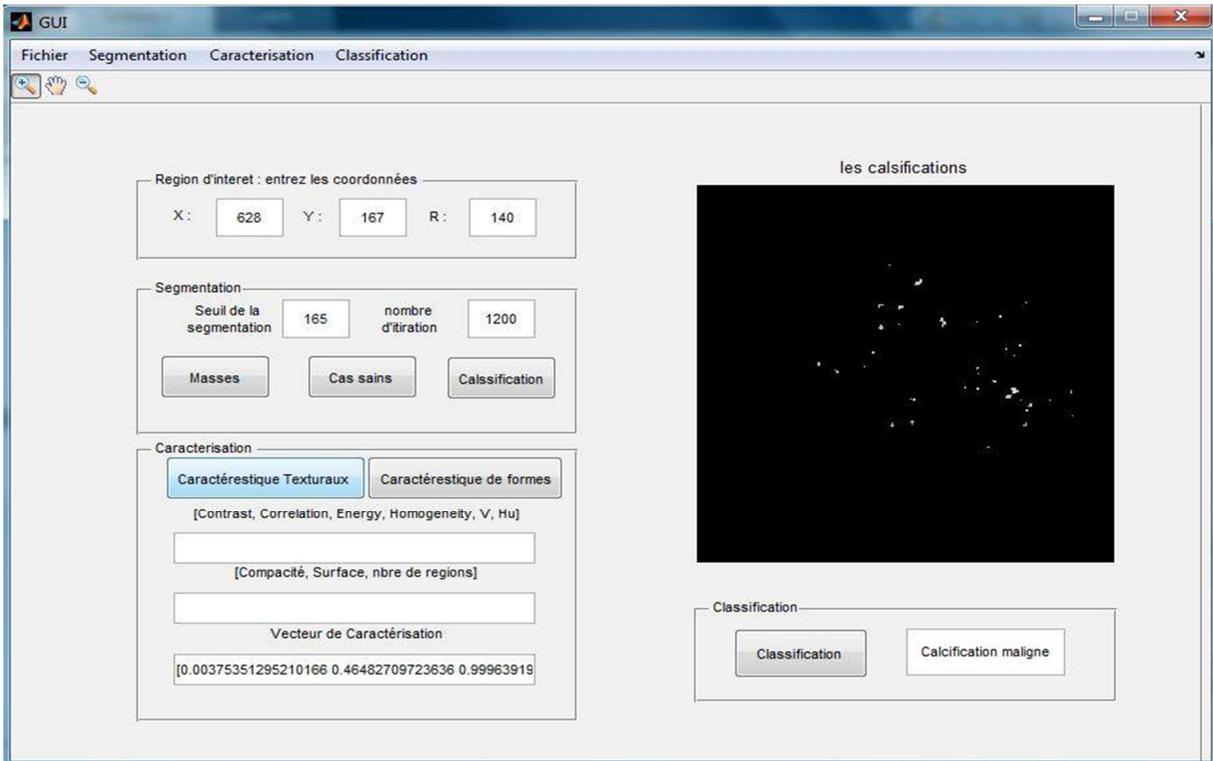


Figure 4.28: classification de calcifications malignes (mdb223)

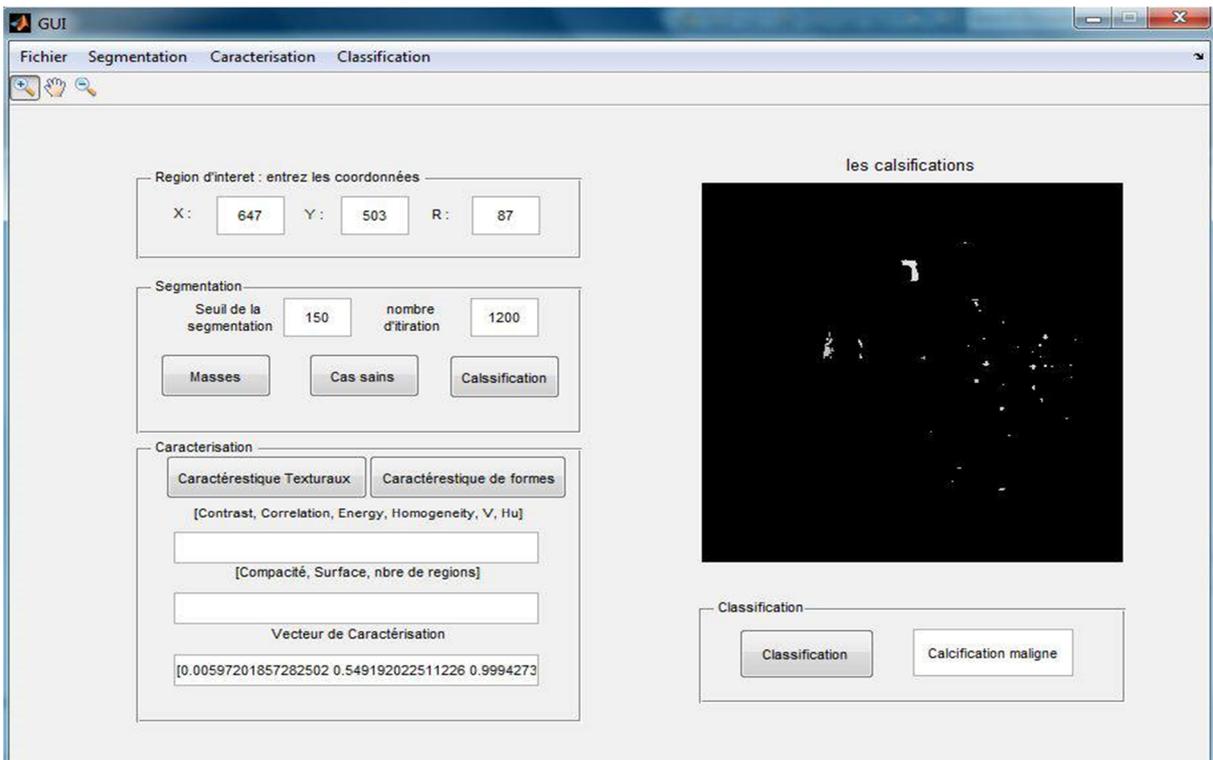


Figure 4.29: classification de calcifications malignes (mdb209)

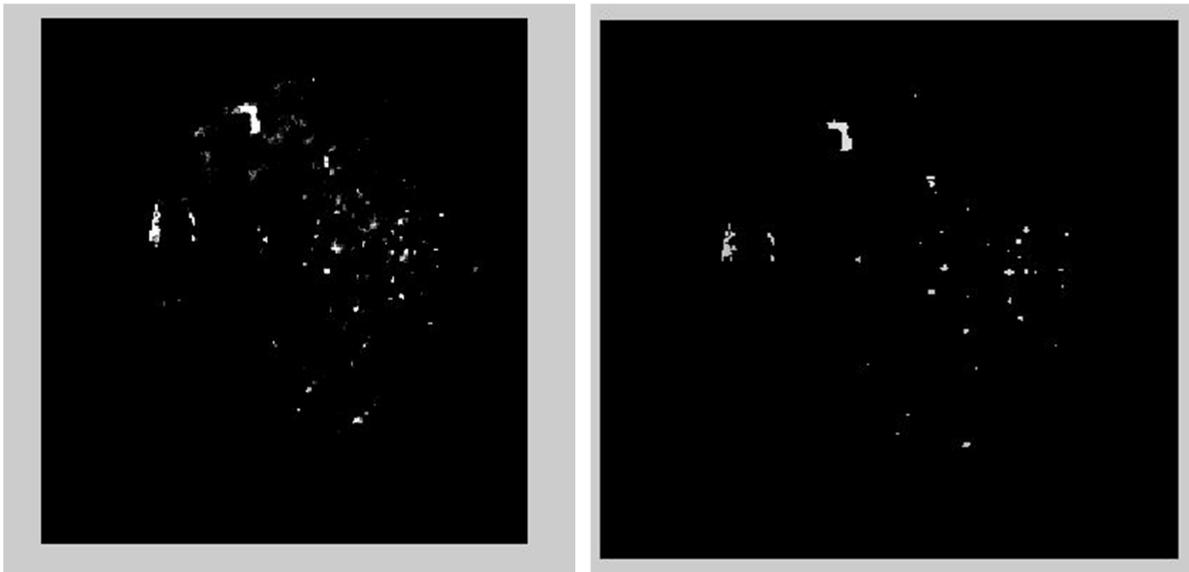


Figure 4.29.a : résultat de l'analyse de l'histogramme Figure 4.29.b : résultat du contour actif

Seuil utilisé= 150, nombre d'itérations=1200, cas de mdb209

Ce choix du nombre d'itérations est très élevé par rapport à celui utilisé par les masses, car dans ce cas, c'est une grande zone qui est prise en compte et, surtout un grand nombre de calcifications.

➤ Cas sains

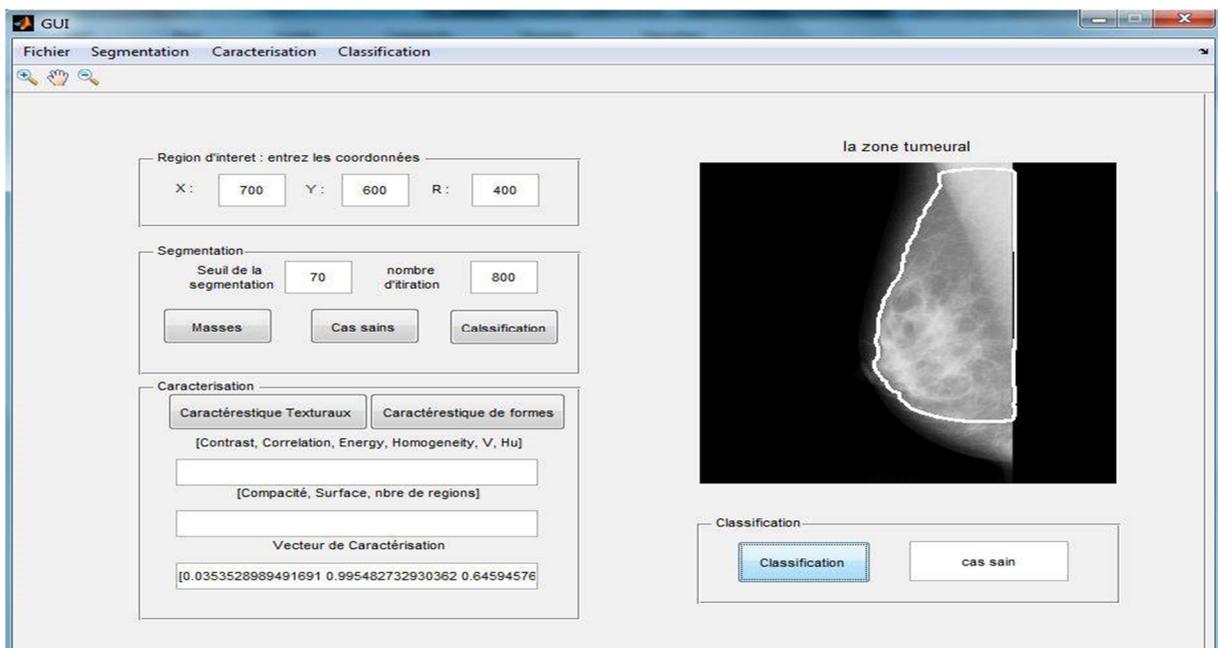


Figure 4.30: classification d'un cas sain (mdb027)

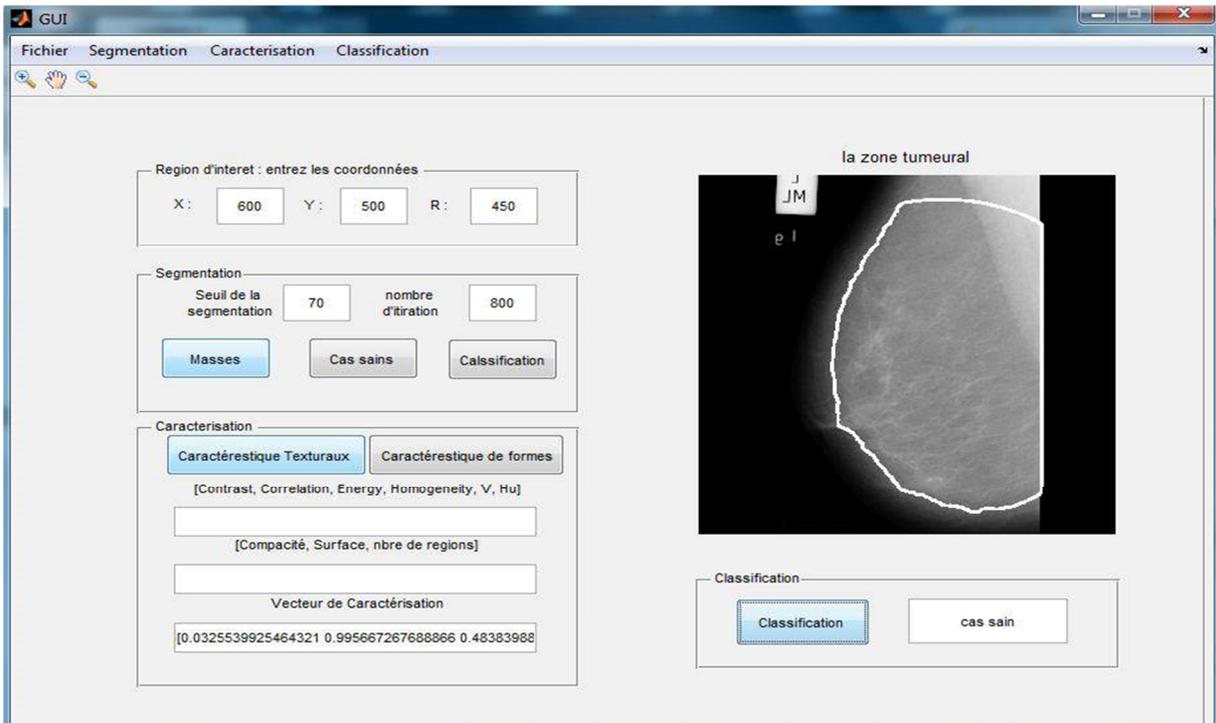


Figure 4.31: classification d'un cas sain (mdb055)



Figure 4.32: classification d'un cas sain (mdb087)



Nous remarquons bien dans ces exemples, que l'extraction des régions d'intérêt, suivant une segmentation coopérative, est obtenue d'une façon plus concise que sans l'hybridation, ce qui a permis une meilleure exploration des descripteurs et du classifieur.

4.4 Conclusion

Pour les différentes pathologies mammaires la détection des tumeurs est toujours difficile à cause de la variété des types de cancers (masses, microcalcifications) et la variabilité des tissus d'une patiente à une autre. Nous constatons que lorsque nous nous consacrons uniquement aux zones suspectes, le système SVNNmam, offre une classification neuronale précise. Ce qui permet d'établir une interprétation sur les différents résultats obtenus, pour les différentes pathologies mammaires étudiées. Dans le même cadre, la segmentation, appliquée sur la zone choisie par le clinicien, indique uniquement les structures pathologiques, suivant l'ajustement des paramètres. Nous arrivons ainsi, à mieux distinguer les différentes formes des masses et des calcifications dans les figures 4.21 ; 4.22, 4.23, 4.24, 4.25, 4.27, 4.28 et 4.29. Ceci nous confirme davantage, que l'analyse de l'image médicale nécessite, l'aide du clinicien pour établir en premier, la zone suspecte.

Cette étude a confirmé la performance du SVNN, dans le domaine de la reconnaissance des pathologies mammaires, suivant un choix judicieux des attributs de caractérisation.



Conclusion générale et perspectives

Dans la vision par ordinateur, la segmentation des images est une tâche qui est très délicate et, pas aisé qui demande une connaissance précise des images et, de leur qualité ainsi qu'une connaissance du terrain si on peut le dire, c'est ainsi que nous avons exploré l'aspect médical durant l'étude de notre mémoire, pour mieux comprendre notre sujet.

Dans ce projet, nous avons étudié et développé, un système de détection et de classification connexionniste de structures pathologiques mammographiques établies par les calcifications, les masses et les cas sains, en vue d'une aide à la décision. Nous nous sommes focalisées sur les travaux rencontrés dans la littérature et les travaux du laboratoire LATSI, pour opter vers une méthode coopérative des contours actifs et des méthodes par analyse d'histogramme, pour prendre en considération les informations des contours et des régions et pouvoir ainsi, délimiter d'une façon concise les régions d'intérêt.

La détection des calcifications, a été performée, en utilisant le chapeau haut de forme, pour l'extraction de toutes les petites structures.

Le processus de la coopération séquentielle réalisé suivant les modèles déformables et les méthodes par analyse d'histogramme, , a prouvé sa robustesse, lorsque nous choisissons les paramètres convenables, à savoir le nombre d'itérations pour les contours actifs et le seuil pour les techniques par analyse d'histogramme. L'initialisation est effectuée, par l'approche région et le raffinement par l'approche contour. Les résultats engendrés avec une variation de paramètres, sont satisfaisants, puisque l'application coopérative sur des images mammographiques, a détecté la forme des lésions convenablement, en tenant compte des concavités et des points anguleux, ce qui n'est pas le cas pour les techniques de segmentation prises séparément.

Il a été montré que l'analyse de forme et de la texture, permet de caractériser les différentes anomalies mammaires rencontrées d'une manière intéressante. Les facteurs de forme,



fournissent des mesures quantitatives de la rugosité de formes. En effet, il a été vu, qu'une masse spinale présente en général, une valeur de compacité plus élevée qu'une masse lisse. La dernière étape de notre système, est la phase de classification d'images mammographiques, suivant le modèle neuronal du perceptron multicouche, associé aux algorithmes génétiques. Notre classifieur SVNN, nous a permis d'obtenir une base d'entraînement comportant des exemples de chaque forme avec, des performances intéressantes. Notre réseau, a pu classier les cinq types de classes avec un taux de reconnaissance de 97%.

Notre système d'aide à la décision, engendre certes, un taux de reconnaissance acceptable, mais des améliorations sont nécessaires pour l'enrichir. Comme l'automatisation de l'initialisation du contour. D'autres techniques peuvent être étudiées, telle que la coopération entre les opérations de la morphologie mathématique, qui tiennent compte de la forme, pour extraire le masque et les ensembles de niveaux, pour faire évoluer le contour. L'entreprise d'une étude de classification clinique par un autre type de réseaux de neurones d'une part et par, les modèles Bayésiens d'autre part, serait envisageable pour comparer les deux approches; dont un certain nombre de travaux théoriques récents permettent de penser qu'elles sont extrêmement voisines. L'emploi d'autres classifieurs en cascade, performerait encore le taux de reconnaissance.

Notre système constitue un outil d'aide à la décision, qui peut donner aux radiologues et aux experts une seconde opinion dans la détection des anomalies.

Bien sur, d'autres options peuvent être explorées, car comme nous le savons, le monde de la recherche n'est pas prêt de s'arrêter.



Bibliographie

[1] : <http://www.ligue-cancer.net/localisation/sein/>

[2] : H. Chekkaf et I. Touil, segmentation des masses dans les images mammographiques, mémoire de Master en Informatique, Université de Tlemcen, 2012

[3]: Eureka Santé, symptômes et évolution du cancer du sein, [http://www.eurekasante.fr/maladie/cancers/cancer des seins](http://www.eurekasante.fr/maladie/cancers/cancer%20des%20seins), Novembre 2012

[4] : B. Séradour, Le dépistage du cancer du sein: un enjeu de santé publique, édition Springer, 2007

[5]: American College of Radiology, ACR BI-RADS - Mammography, Ultrasound and Magnetic Resonance Imaging, edition Reston, 2003

[6]: A.S Morrison, Screening in chronic disease.2nd edition, New York NY, Oxford University Press, 1992

[7]: S.A Feig, Age related accuracy of screening mammography, how should it be measured? Journal of Radiology, 2000; 214, pp 633-640

[8] : J. P. Cocquerez ,S. Philipp, Analyse d'images: filtrage et segmentation, Masson, Paris, 1995

[9]: A. Chanu, Tutorial sur la segmentation en imagerie médicale, <http://www.ai.polymtl.ca/cohen/ELE6812/Tutoriels/Chanu/index.htm>, 2005

[10]: C.-H. Wei*, S.Y Chen, and X. Liu, "Mammogram retrieval on similar mass lesions," Journal of Computer Methods and Programs in Biomedicine, vol. 106, no. 3, pp. 234-248, 2012.

[11]: B. C. Patel, G.R.Sinha, an adaptive K-means clustering algorithm for breast Image segmentation, International Journal of Computer Applications 10(4):35–38, November 2010



[12] : W. Eziddin, segmentation itérative d'images par propagation de connaissances dans le domaine possibiliste: application à la détection de tumeurs en imagerie mammographique thèse de Doctorat, Université de Bretagne, Juin 2012

[13] : S. Bres, J. Jolion, F. Lebourgeois. Traitement et analyse des images numériques, édition Lavoisier, Paris, Octobre 2003

[14] : J. J. Rousselle, les contours actifs, méthode de segmentation : application à l'imagerie médicale, thèse de Doctorat, Université F. Rabelais de Tours, juillet 2003

[15]: J.M. Rendon Mancha, " régions actives morphologiques: application à la vision par ordinateur ", thèse de Doctorat, Université R. Descartes, Paris V, 2002

[16] : M. Meliani, segmentation d'Image par coopération régions-contours, mémoire de Magister en informatique, Ecole Doctorale STIC, Alger, 2010

[17] I.Sebari et D.Chen. HE, les approches de segmentation d'images par coopération régions-contours, Revue de télédétection, vol. 7, n° 1-2-3-4, p. 499-506, 2007

[18] M.Coster et J.L.Cherman, « Traitement d'image par morphologie mathématique application au domaine médical et industriel », Thèse de doctorat, université B. Pascal, 1987

[19] S. Beucher, Segmentation d'images et morphologie mathématique, thèse de doctorat, Ecole Nat. Sup. des Mines de Paris, 1990.

[20] : A.RICK, Représentation de la variabilité dans le traitement d'images flou, application à la mammographie numérique, Thèse de doctorat, Université Paris VI, 1999

[21]: S. Liu , E.J. Delp, Multiresolution detection of stellate lesions in mammograms, IEEE transactions on medical imaging, pp 10-112, 1997.

[22]: F.Z.Reguiég, N.Benblidia, M.Guerti, Morphometric appreciation of mammary breast calcifications around multiresolution analysis by inertia and frequential factors, ACIT2005, Amman, Jordanie, Décembre 2005

[23]: R.M. Rangayyan, Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer , Editor(s) : Jasjit S. Suri; SPIE Press Book, April 2006



- [24]: L. Shen, R. M. Rangayyan, J.E L.Desautels, Detection and classification of mammographic calcifications, International Journal of Pattern Recognition and Artificial Intelligence, Vol 7, N^o 6, pp 1403-1415, June 1993.
- [25]: C.VACHIER, extraction de caractéristiques, segmentation d'image et morphologie mathématique, thèse de doctorat, école des mines de Paris, 1995
- [26]: F. Boudierba, A. Y. Brahim, analyse topologique d'images mammographiques, mémoire d'ingénieur, Département d'électronique, Université S. Dahlab de Blida, Septembre 2009
- [27]: K. F. Cissé, caractérisation texturale des images mammographiques, mathématique, mémoire de Master, Département d'électronique, Université S. Dahlab de Blida, Juillet 2011
- [28]: W. Benhenia, F. Djillali, analyse texturale de cellules tumorales mammaires pour l'aide à l'interprétation, mémoire de Master, Département d'informatique, Université S. Dahlab de Blida, Juillet 2012
- [29]: S.Haouli, analyse texturale et topologique d'images mammographiques, pour l'aide au diagnostic, mémoire d'ingénieur, Département d'électronique, Université S. Dahlab de Blida, Septembre 2012
- [30] S. Herlidou, caractérisation tissulaire en IRM par l'analyse de texture : étude du tissu musculaire et de tumeurs intracrâniennes, Thèse de doctorat, Université de Rennes, 1999
- [31]: I. Claude, analyse de textures en traitement d'images, rapport d'étude, conservatoire botanique national alpin, 2003
- [32] : C.H.Chen, "a study of texture classification using spectral features", Proceedings of the 6th international conference of pattern recognition, Munich 1982
- [33]: M. Hanifi, extraction de caractéristiques de texture pour la classification d'images, thèse de Doctorat, Université de Toulouse, 2009
- [34] : K.Hassan, Exploitation des grandes bases de données d'image sur la grille de calcul, Mémoire de DEA, Université Lumière Lyon 2, France, Juillet 2003
- [35] :R. kachouri, classification multi-modèles des images, thèse de doctorat, université d'Evry-val d'Essonne, 2010
- [36]: M.K.Hu, visual pattern recognition by moment invariants, IRE, transactions on information theory, V. 8 N^o2, pp 179-187, 1962
- [37] : C. Bouveyron, modélisation et classification des données de grande dimension application à l'analyse d'images, thèse de Doctorat, université J. Fourier, 2006



[38] : P. Gagné, Reconnaissance des formes en parallèle utilisant des réseaux de neurones et une mire de réduction de dimensions, thèse de Doctorat, université Laval, Canada, 1999

[39] : G. Dreyfus, M. Samuelides, J.-M. Martinez, M. B. Gordon, F. Badran, S. Thiria, L. Héroult, réseaux de neurones (méthodologies et applications), Eyrolles, 2004

[40]: C. Collet, R. Caloz, Précis de télédétection: volume 3, Traitements numériques d'images, presses de l'université du Québec, 2001

[41] : D.E Goldberg,., *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., New York, NY, USA, (1989)

[42]: L. Ziani, Introduction aux algorithmes génétiques, <http://kijack.perso.neuf.fr/guill.web-/Algogen.html>, 2000

[43]: O. Ludwig, study on non-parametric methods for fast pattern recognition with emphasis on neural networks and cascade classifiers, University of Coimbra, 2012

[44] : J. J. Rousselle, les contours actifs, méthode de segmentation: application à l'imagerie médicale, thèse de Doctorat, Université F. Rabelais de Tours, juillet 2003

[45] Medical Image/Video Engineering Group, MIAS: Mammographic Image Analysis Society, <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>

[46]: R. C. Gonzalez, Digital Image processing using Matlab, Pearson publication, 2005