

الجمهورية الجزائرية الديمقراطية الشعبية

541
17004541

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

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

Département d'Informatique



Mémoire de Fin d'études Master

16/929

Filière : Informatique

Option : Ingénierie des Medias

Thème :

Sélection de caractéristiques pour la classification
des masses mammographiques

Encadré Par :

Mr: Ferkous Chokri

Présenté par :

Sahtel Okba

Benkirat Adel

Juin 2016

Remerciement

En préambule a ce mémoire nous remercions ALLAH qui nous aide et nous donne la patience et le courage durant ces longues années d'étude.

Nous souhaitant adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'a la réussite de cette formidable année universitaire.

Ces remerciements vont tout d'abord au corps professoral et administratif du département d'informatique, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenons a remercier sincèrement Monsieur, Ferkous Chokri, qui, en tant que Directeurs de mémoire, qui est toujours montrés a l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'inspiration, l'aide et le temps qu'ils ont bien voulu nous consacrer et sans qui ce mémoire n'aurait jamais vu le jour.

Enfin, nous adressons nos plus sincères remerciements a tous nos proches et amis, qui nous ont toujours soutenue et encouragée au cours de la réalisation de ce mémoire.

Merci à tous et à toutes

Dédicace

Je dédie ce travail A ...

Ma mère

Mon père

Ma chère sœur

Mes chères amis

A tous les membres de ma famille, petits et grands

Adel

Dédicace

Je dédie ce travail A ...

Ma mère

Mon père

Mes chères sœurs

Mon frère

Mes chères amis

A tous les membres de ma famille, petits et grands

Okba

Résumé

Le cancer du sein est un véritable fléau de nos jours puisqu'il touche une grande population féminine et constitue à lui seul le plus grand pourcentage de mortalité chez les femmes. Cependant la détection de la tumeur à une étape précoce augmentera considérablement les chances de guérison des patientes. Il apparait indispensable de développer de nouvelles méthodes ou de nouveaux protocoles pour le dépistage du cancer du sein.

Dans ce mémoire nous proposons un système pour l'extraction des caractéristiques et la classification des masses mammographiques, ce système s'inspire globalement de l'approche du médecin lors de l'examen radiologique en se basant sur le système d'aide à la rédaction des comptes rendus BI-RADS (Breast Imaging Reporting System and Data System) qui permet de décrire les anomalies rencontrées en mammographie, La segmentation des masses dans notre approche est manuelle car on suppose que la détection est déjà faite. Nous avons validé notre système sur un échantillon de 114 clichés mammographique très subtils provenant de la base DDSM "Digital Database for Screening Mammography", et les résultats obtenus sont très encourageants.

Mots clés :

Mammographie, BI-RADS, Extraction des caractéristiques, La classification, Réseaux de neurones.

Abstract

Breast cancer continues to be a significant public health problem in the world because it affects a large female population, it is also the largest percentage of mortality among women. Early detection is the key to improving breast cancer prognosis. It appears essential to develop new methods or new protocols for the detection of breast cancer.

In this thesis we propose a system for features extraction and classification of mammographic masses, this system is inspired overall by the approach of the doctor during the radiologic examination as it was agreed in BI-RADS (Breast Imaging reporting System and Data System) which makes it possible to describe the abnormalities encountered in mammography. The segmentation of the masses in our approach is manual because we assume that the detection is already made. We present the validation of our method using 114 very subtle mammograms from the Digital Database for Screening Mammography " DDSM" , and the results are very encouraging.

Keywords :

Mammography, Features extraction, BI-RADS Classification, Neural network.

Sommaire

Introduction générale	8
1 Vision par ordinateur et système d'aide au diagnostic	10
1.1 Introduction	10
1.2 La vision par ordinateur	10
1.2.1 Difficulté de la vision par ordinateur	11
1.3 Analyse d'image	11
1.3.1 Analyse de bas niveau d'image	12
1.3.2 Analyse de haut niveau d'image	13
1.4 La reconnaissance de formes	13
1.4.1 Principe des RDF	14
1.4.2 Processus de RDF	14
1.4.3 Domaine d'application	17
1.5 Les systèmes d'aide au diagnostic(SAD)	17
1.5.1 Composition du SAD en mammographie	17
1.6 Conclusion	23
2 Classification des anomalies mammographiques	24
2.1 Introduction	24
2.2 Imagerie mammaire	24
2.2.1 Échographie	24
2.2.2 Imagerie par Résonance Magnétique(IRM)	25
2.2.3 Mammographie	26
2.3 Les anomalies mammographiques	28
2.3.1 Les micro calcifications	28
2.3.2 Les masses	31
2.4 La classification des anomalies mammographiques	33
2.4.1 La classification de LeGal	33
2.4.2 La classification de Lanyi	34
2.4.3 La classification de BI-RADS	34
2.4.4 Synthèse	39
2.5 Conclusion	39
3 Extraction des caractéristiques	40
3.1 Introduction	40
3.2 Les caractéristiques des masses mammographiques	40
3.2.1 Les descripteurs de texture en mammographie	41
3.2.2 Les descripteurs de forme en mammographie	49
3.2.3 Les descripteurs de contour en mammographie	57

3.3	Les caractéristiques humaines	62
3.4	Conclusion	62
4	Conception et réalisation	63
4.1	Introduction	63
4.2	Environnement et outils de développement	63
4.2.1	Plateformes utilisées	63
4.2.2	Langage et environnement	63
4.2.3	La base d'images utilisées	64
4.3	Approche proposée	66
4.3.1	La segmentation	67
4.3.2	Extraction des caractéristiques	67
4.3.3	La sélection des caractéristiques	76
4.3.4	Classification	76
4.4	Réalisation et interface	77
4.5	Test et résultat	79
4.6	Conclusion	80
	Conclusion générale	81
	Bibliographie	81

Table des figures

1.1	Processus de la RDF	14
1.2	Composition d'un système d'aide au diagnostique	18
1.3	Le résultat de segmentation de deux masses circonscrite et spiculée	22
2.1	Exemples d'images échographiques du sein.	25
2.2	Exemple d'IRM mammaire bi-sein	26
2.3	Les composants d'un mammographe	27
2.4	Exemples d'incidences en mammographie : a) Cranio Caudale (CC), b) Médio Latérale Oblique (MLO) et c) Profil interne	28
2.5	Les différents types de micro calcifications	30
2.6	Les différentes distributions des micro calcifications.	31
2.7	Les différentes formes possibles d'une masse.	31
2.8	Les différents contours possibles d'une masse.	32
2.9	Densité mammaire selon le lexique BIRADS.	33
2.10	Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2.	35
2.11	Microcalcifications bénignes : lait calcique : BI-RADS 2.	36
2.12	Masses bien circonscrites discrètement lobulées, sans micro lobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3.	37
2.13	Micro calcifications granulaires irrégulières peu nombreuses : BIRADS 4	37
2.14	Images spiculée sans centre dense : centre prolifératif d'Aschoff.	38
2.15	Masse mal circonscrite aux contours flous et irréguliers : BI-RADS 5.	38
2.16	Micro calcifications vermiculaires, arborescentes, nombreuses et groupées : BI-RADS 5.	38
3.1	La co-occurrence pour un pixel(jaune)	43
3.2	Run Difference Matrix (RDM)	45
3.3	Relation entre complexité du contour et malignité :a)masse circonscrite circulaire, b) circonscrite ovale, c) macro lobulée, d) micro lobulée, e)irrégulière, f) spéculé	49
3.4	Calcul de la mesure de circularité	51
3.5	Calcul de la mesure de rectangularité	52
3.6	La coque convexe et le contour d'une masse mammaire.	53
3.7	Un exemple de caractéristiques 2-6 à l'aide d'une masse en rotation à des angles différents.	54
3.8	Exemple de deux ondelettes l_1 et l_2	59
3.9	Capture des ondelettes avec le taux 10 sur le contour	59
3.10	Modèle schématique du calcul des probabilités de contour de chaque forme d'onde	60
4.1	Les informations importantes dans un fichier ics	65

4.2	Un exemple d'un fichier overlay	66
4.3	Le schéma de notre application	67
4.4	Les caractéristique utilisées dans notre approche	68
4.5	Exemple sur la sélection des points v_i	68
4.6	L'architecture du PMC utilisée.	77
4.7	Interface principale de notre application	78



Liste des tableaux

2.1	La classification de LeGal.	34
3.1	Les avantages et les inconvénients des descripteurs géométriques et spécifiques	58
3.2	Interprétation BI-RADS pour les valeurs de densité du sein	62
4.1	Les statistiques de la base DDSM	79

Liste des abbréviations

ACR	American College of Radiology
AMDI	Indexed Atlas of Digital Mammograms
BE	La boîte englobante d'un objet
BI-RADS	Breast Imaging-Reporting And Data System
DAD	La distribution de distance moyenne
DDSM	Digital Database for Screening Mammography
DGD	La distribution de niveau de gris
DOD	La distribution de différence moyenne
GLCM	Gray Level Co-occurrence Matrix
IBM	International Business Machines
IDE	Integrated Development Environment
IRM	Imagerie par Résonance Magnétique
LDEL	Long Distance Emphasis for Large difference
LDE	Large Difference Emphasis
LJPEG	LOSSLESS JPEG
LRNM	Longueur Radiale Normalisé Modifié
LRN	Longueur Radiale Normalisé
Mcs	Micro Calcifications
MIAS	Mammographic Image Analysis Society
PMC	Perceptron Multi Couches
PMF	La Fonction de Probabilité de Masse
RDM	Run Difference Matrix
RNA	Réseau de Neurone Artificiel
ROI	Region Of Interest
SGBD	Système de Gestion de Base de Donnée
SGLD	Spatial Gray Level Dependency
SMG	Second Moment de DGD
SWT	Standard Widget Toolkit

Introduction générale

Au cours des dernières années, le cancer du sein s'est largement imposé comme un problème majeur de santé dans le monde. C'est le cancer le plus fréquent chez la femme, il représente la première cause de mortalité féminine dans la tranche d'âge de 35 à 55 ans, constituant ainsi un sérieux problème de santé publique, il est responsable de 20% des décès dus aux cancers. En Algérie, selon les statistiques de l'institut national de la santé publique Algérienne (INSP)[1] le cancer du sein représente 30% des cas de cancer recensés, et représente près de 50% des cancers gynécologiques chez la femme, au cours de ces 15 dernières années l'incidence du cancer du sein a été multipliée par 3. Du fait de son diagnostic tardif, il en résulte souvent un traitement lourd, mutilant et coûteux qui s'accompagne d'un taux de mortalité élevé. L'incidence du cancer du sein reste croissante en Algérie et il n'existe toujours pas des programmes de dépistage organisé à l'instar de nos voisins Maghrébins ou Européens.

Dans le but d'éviter des traitements lourds et de réduire la morbidité et la mortalité par le cancer du sein, une détection précoce est nécessaire motivant ainsi des campagnes de dépistage chez les femmes à partir d'un certain âge variant entre 40 et 50 ans.

Le type d'imagerie actuellement utilisée pour cette tâche est la mammographie. La mammographie reste la technique de référence incontournable pour l'exploration du sein, la plus performante en matière de surveillance et de dépistage précoce du cancer du sein. Toutefois, en raison des différents types de fatigue et les facteurs humains, le taux d'échec a été élevé. Il a été observé que les radiologues manquent environ 10% de l'ensemble des lésions cancéreuses.

Lors d'un examen mammographique, le radiologue ne fait que résoudre un problème de vision. C'est à partir de ce point et de récentes techniques de l'analyse et prétraitement d'image, que l'idée d'utiliser la machine pour faciliter le travail du radiologue pour mieux diagnostiquer un cliché mammographique.

Un système d'aide à la décision typique doit comprendre trois étapes importantes qui sont la segmentation, l'extraction des caractéristiques et la classification. La phase d'ex-

traction des caractéristiques nous intéresse particulièrement dans nos recherches.

Dans ce mémoire nous proposons un système qui permet d'extraire à partir d'une masse mammographique un vecteur de caractéristiques qui sera utilisé par la suite dans la phase de classification. La structure du système d'extraction des caractéristiques des masses mammographiques que nous proposons s'inspire globalement de l'approche du médecin lors de l'examen radiologique, la segmentation des masses dans notre approche est manuelle car on suppose que la détection est déjà faite.

Notre mémoire se compose de quatre chapitres :

Chapitre 1 : La vision par ordinateur et les systèmes d'aide au diagnostic

Dans le premier chapitre, l'analyse et l'interprétation de l'image constituent l'outil principal de la vision par ordinateur, on s'intéresse sur les tâches de la vision par ordinateur qui sont divisées en problèmes de bas niveau et haut niveau.

Ensuite, on focalise sur la composition des systèmes d'aide au diagnostic.

Chapitre 2 : La classification des anomalies mammographiques

Dans le deuxième chapitre, Les outils d'imagerie médicale permettant le dépistage et le diagnostic du cancer du sein sont abordées dans l'objectif d'identifier les avantages et les inconvénients de chaque outil. Ensuite, nous détaillons les caractéristiques (morphologie, distribution, forme, contour, ...) des lésions mammaires qui sont les micro-calcifications et les masses, dans l'objectif de différencier entre les cas pathologiques et les cas bénins. Le lexique abordé dans cette caractérisation sert à mieux présenter la classification celle de BIRADS qui est utilisée dans notre travail.

Chapitre 3 : Extraction des caractéristiques

Dans ce chapitre, On commence par aborder les descripteurs de texture les plus fréquents tels que les statistiques de premier ordre, la matrice de co-occurrence, différence matrice (RDM), la transformée de Fourier et l'analyse fractale.

Ensuite, on se focalise sur les différents descripteurs de forme étant donné qu'ils s'avèrent plus performants que ceux de texture dans le cas de lésions de type masse. L'étude de ces descripteurs permet d'identifier les avantages et les inconvénients de chaque méthode, à la fin on se termine par les descripteurs de contour.

Chapitre 4 : Conception et réalisation

Dans ce dernier chapitre nous passons à décrire notre approche que nous avons développée avec une description en détail sur ses composants et les caractéristiques calculées, et la base d'images utilisée.

Chapitre 1

Vision par ordinateur et système d'aide au diagnostic

1.1 Introduction

L'analyse et l'interprétation de l'image constituent l'outil principal de la vision par ordinateur, en abordant des problèmes dont les humains résolvent à leur insu. L'objectif de la vision par ordinateur est de développer des machines qui reconstruisent et interprètent un environnement tridimensionnel basé sur des mesures d'énergie radiante[6].

Dans ce chapitre, on focalise sur les tâches de la vision par ordinateur qui sont divisées en problèmes de bas niveau et haut niveau ensuite, on va détailler le processus de la reconnaissance de formes, et à la fin on se concentre sur le système d'aide au diagnostique et ses composants.

1.2 La vision par ordinateur

La vision permet aux humains de percevoir et de comprendre le monde qui les entoure, la vision par ordinateur vise à reproduire l'effet de la vision humaine en percevant électroniquement une image [53].

En contre partie, donner à l'ordinateur la possibilité de voir n'est pas une tâche facile, en effet, nous vivons dans un monde à trois dimensions (3D), et quand les ordinateurs tentent d'analyser des objets dans l'espace 3D, les capteurs visuels disponibles comme les caméras de télévision, donnent généralement des images bidimensionnelles (2D), et cette projection en un nombre inférieur de dimensions encourt une énorme perte d'informations. La majorité des techniques de vision par ordinateur se basent sur les résultats issus des méthodes de mathématiques, reconnaissance de formes, intelligence artificielle,

psycho-physiologie, informatique, électronique, et d'autres disciplines scientifiques.

La création ou l'utilisation des algorithmes ainsi que des dispositifs d'analyse d'image numérique, est conditionnée par la prise en compte des principes de la perception humaine de l'image. Lorsqu'un humain analyse une image, il exprime l'information contenu au sein de cette dernière en utilisant des descripteurs faciles à percevoir, c'est des descripteurs psychophysiques tels que la couleur, le contraste, les contours, la forme, la texture...etc.

L'analyse de la texture est un domaine important pour la vision par ordinateur. En effet, la majorité des surfaces naturelles présentent une texture et un système performant de vision doit être capable de distinguer la texture de l'environnement pour lequel il a été conçu.

1.2.1 Difficulté de la vision par ordinateur

Beaucoup de raisons font que la vision par ordinateur soit assez complexe, nous citons [53] :

- La perte d'informations suite au passage du 3D à 2D, en effet, c'est un phénomène qui se produit durant la capture d'image par un appareil photo.
- Le bruit est présent de façon inhérente dans chaque mesure au sein du monde réel. Son existence appelle aux outils mathématiques pouvant l'éliminer. En contre partie, des outils plus complexes rendent l'analyse d'image beaucoup plus compliquée par rapport aux méthodes standards.
- Les quantités des images et des séquences vidéo sont énormes. Par exemple une feuille de papier A4 numérisée monochromatique à 300 points par pouce (ppp) à 8 bits par pixel correspond à 8,5 MB. Citons un autre exemple d'une vidéo RGB de 24 bits couleur 512x768 pixels, 25 images par seconde, rend un flux de données de 225 MB par seconde. Cela nécessite la conception d'un traitement simple et pertinent, dans le cas contraire, il serait difficile d'obtenir des performances en temps réel, c'est à dire, pour traiter 25 ou 30 images par secondes.
- La mesure d'intensité lumineuse ou du rayonnement dépend de l'éclairement énergétique c'est-à-dire le type de la source de lumière, la position de l'observateur etc. Un lien direct entre l'apparence des objets dans les scènes et leur interprétation est distingué.

1.3 Analyse d'image

Une image numérique est constituée d'une grille rectangulaire d'échantillonnage dont les constituants sont des pixels portant des informations sur l'intensité lumineuse des dif-

férents lieux au sein de l'image. De ce fait, Une image numérique est représentée par une matrice bidimensionnelle, dont les éléments sont des nombres naturels correspondant à des niveaux de quantification dans l'échelle de l'intensité lumineuse.

L'analyse d'image de bas niveau et la vision par ordinateur de haut niveau diffèrent dans les données utilisées. En effet, les données de bas niveau sont constitués d'images d'origine représentées par des matrices composées de valeurs d'intensité lumineuse, tandis que les données de haut niveau sont originaires des images aussi, en revanche, seulement les données qui sont pertinentes et qui se rapportent à des objectifs de haut niveau sont extraites, ce qui réduit considérablement la quantité de données.

1.3.1 Analyse de bas niveau d'image

Les techniques de bas niveau de la vision par ordinateur représentent la base du traitement numérique de l'image[53], ils utilisent très peu de connaissances sur le contenu des images. Tout d'abord une image d'entrée est capturée par une caméra de télévision en (2D) et numérisée, étant décrite par une fonction d'image $f(x, y)$ dont la valeur est en général l'intensité lumineuse en fonction de deux paramètres (x, y) représentant les coordonnées de l'emplacement dans l'image. Puis l'étape du traitement qui consiste en les opérations effectuées sur les images au plus bas niveau d'abstraction comme l'élimination du bruit au sein de l'image, voir plus, l'amélioration de certains descripteurs des objets jugés pertinents pour interpréter l'image. L'entrée et la sortie sont des images d'intensité. Ces images iconiques sont généralement de la même nature que les données originales capturées par le capteur, le traitement ne fait pas augmenter les informations du contenu de l'image, bien au contraire, il diminue généralement les informations contenues au sein de l'image. Du point de vue de la théorie de l'information, le meilleur traitement est de ne pas faire un traitement.

La meilleure façon d'éviter l'élaboration d'un traitement est de se focaliser sur une acquisition d'images de haute qualité. Toutefois, le traitement est très utile dans une variété de situations, car il permet de supprimer les informations qui ne sont pas pertinentes pour l'analyse de l'image. Par conséquent, le but du traitement est l'amélioration des données d'image en supprimant les indésirables distorsions et en améliorant certaines caractéristiques importantes de l'image pour l'analyse ultérieure de l'image, bien que les transformations géométriques d'images comme la rotation et le redimensionnement sont également classés comme des méthodes de traitement.

Sachant que les pixels voisins correspondant à un objet donné dans des images réelles ont essentiellement la même valeur d'intensité lumineuse, de sorte que si un pixel déformé

peut être capté à partir de l'image, il peut généralement être restauré comme une valeur moyenne des pixels voisins, ceci est un exemple d'un traitement que nous pourrions effectuer sur une image. La segmentation d'image est la prochaine étape du processus d'analyse d'image, dans laquelle l'ordinateur tente de séparer les objets de l'arrière plan de l'image. Nous distinguons une segmentation totale et une segmentation partielle : la segmentation totale n'est possible que pour des tâches très simples, un exemple étant la reconnaissance d'objets non jointifs sombres du fond clair. Dans des cas de problèmes plus compliqués, les techniques d'analyse d'image de bas niveau gèrent les tâches de segmentation partielle, dans laquelle seuls les indices qui aideront l'analyse ultérieure de haut niveau sont extraits. La description d'objet dans une image totalement segmentée est également comprise dans le cadre d'analyse de bas niveau d'image.

1.3.2 Analyse de haut niveau d'image

L'analyse de haut niveau[53] s'appuie sur des connaissances relatives au contenu de l'image, par exemple, taille de l'objet, sa forme et les relations mutuelles entre les objets dans l'image ; ces données de haut niveau sont généralement exprimées sous une forme symbolique. Les méthodes d'intelligence artificielle sont largement applicables aussi, en effet, la vision par ordinateur de haut niveau tente d'imiter la cognition humaine et la capacité à prendre des décisions en fonction de l'information contenue dans l'image. La vision de haut niveau commence par une certaine forme du modèle formel du monde, puis la réalité perçue sous la forme d'images numériques est comparée au modèle ; l'ordinateur passe en analyse d'image à bas niveau pour trouver des informations nécessaires afin de mettre à jour le modèle.

1.4 La reconnaissance de formes

La reconnaissance de formes(RDF) est un ensemble de techniques et méthodes visant à identifier des motifs à partir des données brutes afin de prendre une décision dépendante de la catégorie attribué à ce motif, elle se considère comme une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques. Le mot forme ne s'agit pas que de forme géométrique. Les formes à reconnaître peuvent être de natures très variés.

Il peut s'agir de contenu visuel(code barre, visage, empreinte digitale) ou sonore(reconnaissance de parole), d'images médicales multi spectrales(images satellitaires) ou rayon X, IRM...) ou bien d'autres. La RDF s'intéresse a la conception et a la réalisation de systèmes (matériels et logiciels) capables de percevoir, et dans une certaine mesure, d'interpréter des signaux captés dans le monde physique[16].

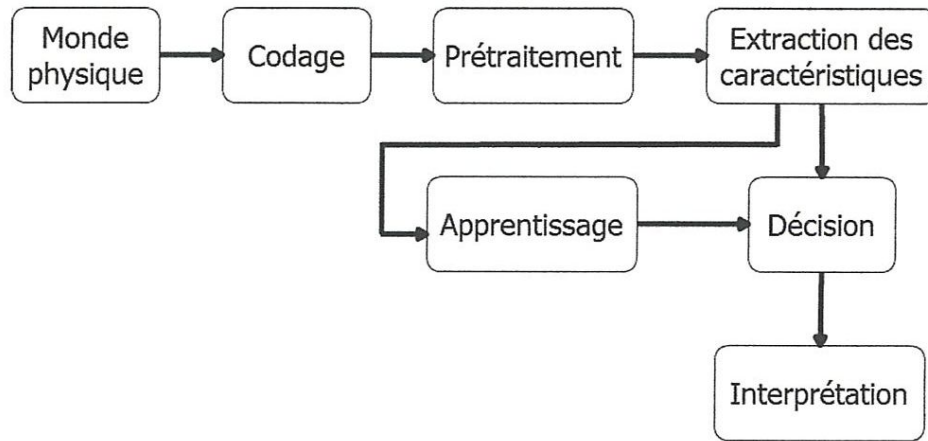


FIGURE 1.1 – *Processus de la RDF*[28]

1.4.1 Principe des RDF

La RDF est un des nombreux aspects de l'intelligence artificielle. A partir d'un ensemble de données ou d'informations apprises, elle offre la possibilité d'interpréter toute nouvelle observation(ou forme). Les observations déjà connues sont regroupées en classes, constituant des prototypes auxquels la nouvelle observation est comparée pour être identifiée. Les algorithmes utilisés permettent donc de classer des observations dont les propriétés ont varié par rapport à une observation type.[28].

1.4.2 Processus de RDF

Les informations issues du monde réel(via le capteur) et fournies au système de RDF sont généralement trop volumineuses et peu pertinentes. Le processus de RDF consiste à la réduction progressive et sélective de l'information, les étapes de traitement d'un processus de RDF(voir la figure 1.1)[48].

1.4.2.1 Monde physique

Le monde physique qui nous entoure est considéré comme un espace analogique de dimension n appelé l'espace de formes F . C'est celui qui est présenté dans sa forme la plus primaire, c'est-à-dire dont nous devons déterminer les caractéristiques les plus apparentes avant l'étape du codage[36].

1.4.2.2 Codage

Est une opération qui consiste en une conversion numérique du monde physique continu vers un monde discret.

1.4.2.3 Prétraitement

Il permet de sélectionner, dans l'espace de représentation, l'information nécessaire à l'application. Cette sélection passe par l'élimination du bruit dû aux conditions d'acquisition, par la normalisation des données ainsi par la suppression de redondance.

1.4.2.4 Extraction des caractéristiques

L'objectif de l'extraction et de la sélection de caractéristiques est d'identifier les caractéristiques importantes pour la discrimination entre classes. Après avoir choisi le meilleur ensemble de caractéristiques, il s'agit de réduire la dimensionnalité de l'ensemble des caractéristiques en trouvant un nouvel ensemble, plus petit que l'ensemble original, qui néanmoins, contient la plupart de l'information.

Quelques exemples sont présentés :

- En reconnaissance des caractères, les caractéristiques utilisables peuvent venir de la densité des points, des moments, des lieux caractéristiques, des transformées mathématiques(Fourier, Walsh, Hadamard...), elles peuvent également venir des squelettes ou des contours.
- Dans des applications liées à l'analyse de texture telles que télédétection et analyse des scènes, les caractéristiques utilisables peuvent venir de la matrice de co-occurrence, des descripteurs de Fourier, du spectre de puissance , des moments, aussi bien que de diverses primitives structurelles.
- Dans l'analyse et la reconnaissance de formes d'ondes telles que le signal sismique, l'EEG et l'ECG, la parole aussi bien que les images de formes courbes, les caractéristiques utilisables peuvent venir du spectre de puissance, de fonctions d'approximation, des zero crossing, et de plusieurs types de segments de traits structurels[38].

1.4.2.5 Apprentissage

L'apprentissage tente de définir des classes de décision ou d'appartenance. Son rôle est de déclarer la décision à l'aide de connaissance a priori sur les formes, à partir de critères spécifiques aux formes. Il existe 2 types d'apprentissages(supervisé, non supervisé)[59].

a) Apprentissage supervisé : Les exemples d'apprentissage sont étiquetés afin d'identifier la classe à laquelle ils appartiennent. Le but de l'algorithme de classification est de classer correctement les nouveaux exemples dans les classes définies dans la phase d'apprentissage. Les méthodes d'apprentissage supervisé se construisent à partir de la base d'apprentissage, des classifieurs, ou fonctions de classement[43].

Si l'utilisateur possède suffisamment d'informations sur la population à étudier, il peut effectuer une classification supervisée. Cette catégorie suppose avoir un groupe d'individus de chaque classe, dont on connaît leur appartenance. Ces individus forment des échantillons "d'apprentissage" . Ils sont utilisés pour entraîner le classifieur. D'autres échantillons, dits "de test" , servent à valider la classification en évaluant sa pertinence à travers le taux d'individus bien classés. Il existe plusieurs méthodes de classification supervisées. Les méthodes les plus réputées sont l'analyse discriminante linéaire, la régression logistique, les réseaux de neurones.

Les réseaux de neurones artificiels, connus par l'acronyme anglais ANN(Artificial Neural Networks), sont largement utilisés pour les problèmes de classification. Ils reposent sur la théorie des perceptrons. Un ANN est composé de plusieurs neurones répartis sur une couche d'entrée (désignant les descripteurs), une couche de sortie (désignant le résultat de classification) et un nombre de couches cachées. Toutefois, l'inconvénient de cette méthode est le choix du nombre de couches cachées et du nombre de neurones dans chaque couche. Ainsi, l'utilisateur est amené à faire des essais avec différentes combinaisons du nombre de couches et de neurones afin d'aboutir au réseau de neurones le plus adapté à son type d'application.

b) Apprentissage non supervisé : L'algorithme d'apprentissage cherche à trouver des régularités dans une collection d'exemples, puisque dans ce type d'apprentissage on ne connaît pas la classe à laquelle les exemples d'apprentissage appartiennent. Une technique employée consiste à implémenter des algorithmes pour rapprocher les exemples les plus similaires et éloigner ceux qui ont le moins de caractéristiques communes[43].

Ces techniques sont utilisées lorsque l'identité des classes n'est pas connue. Cela résulte d'un manque d'information de la population à étudier. La classification non-supervisée , dite automatique , ou groupement connue par(clustering en anglais) consiste à déterminer les différentes classes naturellement sans aucune connaissance préalable.

Parmi les méthodes de classification non supervisées la méthode la plus communément utilisée est celle de l'algorithme K-moyennes également appelée algorithme des nuées dynamiques (en anglais k-means)[31]. L'algorithme fonctionne en précisant le nombre K de classes (clusters) attendues (K étant fixé par l'utilisateur). Il calcule la distance intra-classe et refixe les centres de classe selon les valeurs de distance. Les inconvénients de cette méthode sont premièrement la nécessité de fixer le nombre de classes avant de commencer la classification. Deuxièmement, cette méthode est très sensible à la répartition initiale des données.

1.4.2.6 Décision

La décision ou classement est l'étape proprement dite la reconnaissance son rôle est de classifier la forme ciblée à partir de l'apprentissage réalisée. Pour la décision et pour l'apprentissage, les critères utilisés sont habituellement les mêmes.

1.4.3 Domaine d'application

- Robotique/industrie : assemblage(reconnaissance de pièces), contrôle de qualité, véhicule autonome.
- Télédétection : météo, identification et suivi des cultures, des forêts, des réserves d'eau, cartographie.
- Application militaire : guidage de missile, reconnaissanceaérienne et espionage.
- Bureautique : reconnaissance de texte par ordinateur, l'analyse de documents.
- Sécurité : identification des empreintes digitales, et reconnaissance de visage.
- Médecine : échographie, IRM(Imagerie par Résonance Magnétique) pour détecter des tumeurs, cellules cancéreuses ou autres maladie, et analyse d'images médicales.

1.5 Les systèmes d'aide au diagnostique(SAD)

Un SAD dédiée à l'imagerie médicale est un programme numérique qui analyse des clichés médicaux et consiste à assister le clinicien dans la tâche parfois délicate d'interprétation et de classement des images médicales.

1.5.1 Composition du SAD en mammographie

En pratique, le SAD est une suite de phases qui doivent être exécutées l'une après l'autre, depuis l'acquisition de l'image jusqu'à la prise de décision. Certaines de ces phases sont souvent étroitement liées et indissociables. Les étapes de traitement d'une image mammographique peuvent se résumer en :

- une étape de prétraitement qui sert à améliorer la qualité de l'image avant toutes manipulations.
 - une étape de segmentation qui permet de détecter la lésion à étudier.
 - une étape de description qui a pour but de caractériser les lésions à travers des formulations mathématiques.
 - une étape de classification et de prise de décision en utilisant un classifieur adéquat.
- Ces différentes étapes sont résumées dans le diagramme représenté dans la figure 1.2.

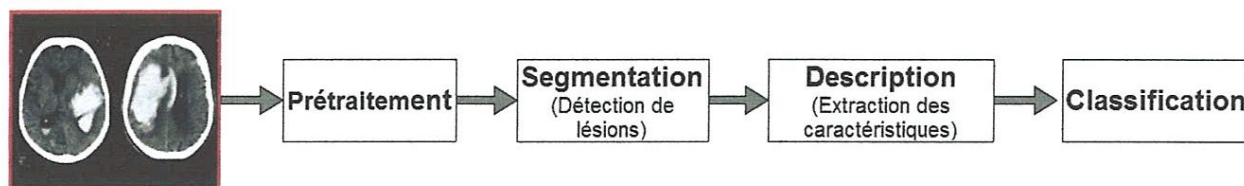


FIGURE 1.2 – Composition d'un système d'aide au diagnostic.

1.4.2.1 Prétraitement

Le cancer du sein (comme tous les cancers d'une manière générale) doit être détecté dans sa première phase pour maximiser les chances de survie. Sauf que dans cette phase, il s'avère très difficile de repérer, à l'œil nu, la pathologie dans le tissu mammaire environnant sans avoir recours à un prétraitement spécifique de l'image acquise. D'où, l'objectif principal de cette étape est d'augmenter le contraste entre la lésion mammaire (que ce soit masse ou micro calcification) et le reste de l'image pour faciliter les traitements ultérieurs. Sachant que dans le cas où une région d'intérêt diffère en luminance de moins de 2% du reste de l'image, elle demeure indiscernable à l'œil nu [14]. Le prétraitement des images mammographiques est connu sous le nom de rehaussement ou d'amélioration du contraste.

Le problème majeur des algorithmes de rehaussement du contraste réside dans le fait que certaines régions peuvent ne pas être rehaussées convenablement alors que d'autres peuvent être l'objet d'un excès de rehaussement. Un manque de rehaussement du contraste peut causer des faux négatifs (FN). En effet, plusieurs détails concernant la lésion peuvent être négligés. Dans ce cas, certaines lésions peuvent ne pas être détectées et par la suite non diagnostiquées. Ce qui ne répond pas à l'objectif principal de la détection précoce d'un cancer. Un excès de renforcement du contraste peut causer des faux positifs (FP). Dans ce cas, plusieurs détails inexistantes réellement peuvent s'ajouter à la lésion. D'où, certaines régions normales du tissu mammaire peuvent être considérées comme lésions ce qui va engendrer des biopsies inutiles.

Les techniques traditionnelles d'amélioration du contraste ont été appliquées à la mammographie pendant plus de trois décennies. La méthode du masque flou a été utilisée par Chan et al[9]. Des études concernant les caractéristiques opérationnelles du récepteur ont montré que la méthode du masque flou parvient à améliorer la détection des micro calcifications. Cependant, cette méthode a causé l'augmentation du bruit dans l'image.

Une approche couramment utilisée pour l'amélioration du contraste est la modification globale de l'histogramme. Cette méthode consiste à réaffecter les valeurs d'intensité des pixels afin de rendre la nouvelle répartition des intensités plus uniforme. Ceci peut être

réalisé par l'égalisation d'histogramme ou par l'étirement d'histogramme [45]. Toutefois, ces transformations ont l'inconvénient de faire ressortir le bruit dans les images déjà bruitées.

Certaines méthodes de rehaussement d'images mammographiques causent une amplification du bruit (la méthode du masque flou, la modification globale de l'histogramme) et par la suite la distorsion des lésions (qui présentent l'information la plus importante à préserver dans de telles images). Les radiologues préfèrent que les images rehaussées maintiennent la même apparence que les images originelles ce qui limite les compétences des techniques de rehaussement. De plus, en se basant sur les résultats fournis par [52], le rehaussement du contraste dans notre cas de traitement des masses mammaires ne s'avère pas nécessaire.

1.4.2.2 Segmentation

La segmentation est souvent considérée comme l'étape initiale dans un système de diagnostic assisté par ordinateur (DAOx) surtout si on fait abstraction de l'étape de pré-traitement qui, d'après la section précédente, n'est pas indispensable dans le cas de traitement des masses. La phase de segmentation est très importante puis que les traitements ultérieurs (description et classification) sont fortement liés au résultat de segmentation. En effet, une bonne détection du contour de la lésion engendre une description fidèle à ses caractéristiques. Ainsi, on peut garantir une classification minimisant le taux des faux positifs et maximisant le taux des vrais négatifs.

Cette étape consiste à détecter ou bien la totalité du sein à partir du fond de l'image ou bien un type d'anomalie bien spécifique comme les micro calcifications et les masses. Il a été démontré que la détection des masses est plus difficile que la détection des micro calcifications. En effet, les masses peuvent être masquées partiellement par le tissu mammaire. Par ailleurs, il y a une grande variabilité dans ces lésions (forme, contour, dimension).

Il existe de nombreuses techniques de segmentation qui ne cessent de se multiplier et de se diversifier. Ces méthodes sont couramment classées en trois catégories : les approches basées pixel, les approches basées contour et les approches basées région. Les approches basées sur les pixels sont généralement fondées sur l'étude des histogrammes de l'image par seuillage, clustering ou clustering flou. Les approches contour abordent la segmentation en tant que recherche de frontières entre les objets (les anomalies) et le fond. Elles consistent à identifier les transitions d'intensité de pixel entre les régions pour définir les bords des anomalies cherchées. Les approches région consistent à partitionner l'image en

régions distinctes vérifiant une certaine homogénéité. Ces méthodes permettent d'aboutir directement à une partition de l'image étant donné que chaque pixel est affecté à une région unique. On détaille dans ce qui suit les trois approches cités précédemment.

a) Les méthodes de seuillage : Le seuillage a pour objectif de segmenter une image en deux ou plusieurs classes. Cette opération consiste à effectuer une partition de l'histogramme en niveaux de gris en utilisant un ou plusieurs seuils. Chaque pic de l'histogramme correspond à une classe. En effet, cette méthode n'est efficace que si l'histogramme contient réellement des pics séparés. Les méthodes de seuillage ont été largement utilisées pour la segmentation de masses mammaires. Par exemple, les auteurs dans [33] ont utilisé différentes valeurs de seuils en niveau de gris qui dépendent du type de tissu mammaire et ceci en se basant sur une analyse de l'histogramme. Plus récemment, Mudigonda et al. [35] ont utilisé un seuillage multiniveaux pour détecter des contours fermés. L'inconvénient majeur de cette approche est le fait de considérer que les masses ont une densité uniforme par rapport au fond de l'image ce qui n'est pas toujours vérifié.

Cependant, l'inconvénient majeur de ces différentes techniques de seuillage est le choix du seuil ou de l'intervalle de seuillage. En effet, avec un intervalle trop large, on obtient des faux positifs. Dans ce cas, l'image seuillée contient des pixels qui ne font pas partie des objets d'intérêt. Il s'agit généralement de bruit ou de pixels (zones) qui ont un niveau de gris proche de celui des objets recherchés. Avec un intervalle trop étroit, on obtient des faux négatifs. Certains objets d'intérêt ou bien des parties de ces objets n'apparaissent pas dans l'image seuillée

b) Les méthodes basées régions : Les méthodes basées région regroupent des pixels ayant des propriétés communes, ainsi ces méthodes donnent directement une partition de l'image. Les méthodes basées région les plus populaires sont la méthode de croissance de régions (region growing) [8] et la méthode de division et fusion de régions (split and merge) [47].

La méthode de croissance de régions est basée sur la propagation d'un point germe initial (seed pixel) selon un critère d'homogénéité spécifique et sur la croissance itérative de la région. Cette méthode comporte en général deux étapes, une étape d'initialisation où l'on trouve les germes des régions les plus représentatives de l'image. Ces germes s'accroissent dans la deuxième phase en respectant des contraintes d'homogénéité, de forme géométrique ou de taille. Les algorithmes de croissance de région ont été largement utilisés pour la segmentation des masses. Toutefois, l'inconvénient majeur de cette approche est l'intervention humaine pour la segmentation de chaque objet.

Une alternative des méthodes de croissance de région est la segmentation par division et fusion de régions (split and merge). L'algorithme "division et fusion" a été proposé par Horowitz et Pavlidis en 1974 [21]. Le processus est décomposé en deux étapes. Dans la première étape, appelée division, on analyse individuellement chaque région de l'image. Si celle-ci ne vérifie pas le critère d'homogénéité, alors on divise cette région en blocs (le plus généralement en 4 quadrants) et l'on réitère le processus sur chaque sous-région.

Cette approche n'a pas été suffisamment exploitée dans le cadre de la segmentation des masses mammaires. La principale difficulté de cette approche réside dans le parcours de l'ensemble de tous les couples de régions voisines et la fixation du critère d'arrêt de la méthode.

c) Les méthodes basées contours : Les approches les plus anciennes de détection de contour font intervenir la notion de dérivation afin de détecter les transitions entre les objets et le fond de l'image qui correspondent à des zones de fortes variations d'intensité. Les méthodes dérivatives utilisées se basent sur des opérateurs tels que Roberts, Sobel, Prewitt et Canny. Une fois la norme et la direction du gradient sont calculées en chaque pixel de l'image, ces méthodes extraient des contours d'un seul pixel d'épaisseur en sélectionnant les maxima locaux des normes des gradients. En mammographie, on est souvent confronté à la présence de bruit (d'acquisition), de textures fines ou de frontières pas très nettes, d'où les transitions détectées ne correspondent pas forcément aux contours réels. Il est alors souvent nécessaire d'appliquer un traitement en aval afin d'écarter les transitions dues aux bruits. De plus, les contours extraits sont généralement discontinus et peu précis. Il faut donc, utiliser des techniques de reconstruction de contours par interpolation ou connaître a priori la forme de l'objet recherché afin de connecter les points du contour.

Dans le cas des approches région, plusieurs travaux [42] se basent sur le fait que les intensités sont statistiquement homogènes dans chaque région à segmenter. Or l'inhomogénéité de l'intensité se produit souvent dans les images médicales comme c'est le cas des images mammographiques. Le modèle du contour actif basé région récemment proposé par [30] est capable de segmenter des images ayant diverses intensités inhomogènes. Par ailleurs, il parvient à fournir un bon résultat de segmentation dans le cas d'objets à contours mal définis ou masqués (ce qui est souvent le cas des masses mammaires). En utilisant le terme de régularisation proposé par [30], la régularité de la fonction "level set" est intrinsèquement préservée ce qui garantit la précision du calcul et évite les procédures coûteuses de réinitialisation.

On présente dans la figure 1.3, le résultat de segmentation d'une masse circonscrite (première ligne) et d'une masse spiculée (deuxième ligne) en utilisant l'algorithme proposé

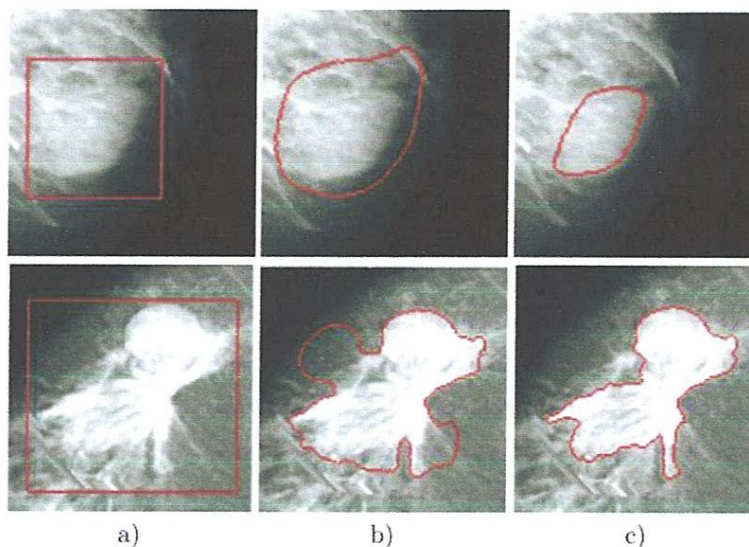


FIGURE 1.3 – Le résultat de segmentation de deux masses circonscrite et spiculée[30].

par[30]. Les figures 1.3.a) montrent la superposition du contour initial sur la région d'intérêt. Les figures 1.3.b) montrent la propagation du contour et les figures 1.3.c) présentent la convergence vers le contour de la lésion.

1.4.2.3 Description

L'être humain reçoit en permanence des informations très diverses et très complexes par l'intermédiaire de ses cinq sens. En dépit de l'abondance de ces informations, le cerveau humain est capable de restituer chaque objet observé et de lui attribuer une représentation cohérente appelée "description humaine".

Dans le domaine du traitement de l'image, la description est l'étape qui cherche à reproduire le même processus d'analyse et d'interprétation. En effet, la description a pour but d'extraire les caractéristiques qui décrivent au mieux et de façon quantitative ou qualitative les objets présents dans l'image. Elle transforme les informations de bas niveau issues de la phase d'acquisition (après probablement prétraitement et segmentation) en informations de haut niveau de telle sorte que les formes et les structures soient décrites de façon analytique. De manière générale, plus la description effectuée est proche de "la description humaine", plus elle est considérée comme robuste et fidèle à l'image initiale. Les méthodes de description d'images sont variables et dépendent de l'objectif visé (description globale, description locale) et du type d'image à analyser (image binaire, image en niveau de gris, image couleur). En littérature, la description d'images est assurée en utilisant la couleur, la texture et/ou la forme. Dans le chapitre 3 on va entamer avec détails les descripteurs de forme et texture.

1.4.2.4 Classification

La classification est considérée comme la dernière étape dans un système d'aide au diagnostic. Elle exploite le résultat de description pour pouvoir décider de la nature pathologique de la masse.

La notion de classification signifie l'affectation d'une étiquette à des échantillons d'une base de données en utilisant un certain nombre de caractéristiques. Ces caractéristiques doivent bien évidemment être capables d'identifier chaque échantillon. En traitement d'images, l'échantillon peut désigner un pixel, une zone dans l'image, un objet représenté dans l'image ou l'image elle-même. Selon l'application, le but de la classification est soit de :

- classer les pixels de l'image en différentes zones. Dans ce cas, le problème de classification revient à un problème de segmentation d'images en différents objets .
A titre d'exemple, on peut classer les différentes zones d'une image mammographique en lésion ou non lésion.
- classer l'image ou les objets de l'image selon différentes catégories. A titre d'exemple, on peut classer les masses qui se trouvent dans les images mammographiques en malignes ou bénignes.

On peut distinguer deux catégories de méthodes de classification : les classifications non-supervisées et celles supervisées qui sont présentées précédement dans la section(1.4.2.5)

1.6 Conclusion

L'analyse d'image et la vision par ordinateur présentent un champ de recherche actif s'évoluant de plus en plus, et l'interprétation de l'image constitue l'outil principal de la vision par ordinateur.

Une étude plus ou moins approfondie concernant les différentes étapes d'un système d'aide au diagnostique a été menée dans ce chapitre. Une telle étude est nécessaire afin de mieux comprendre cette chaîne et de faire des choix appropriés pour le traitement des masses.

Chapitre 2

Classification des anomalies mammographiques

2.1 Introduction

La présentation de l'approche médicale de l'imagerie mammaire est indispensable puisqu'elle nous permet d'acquérir un savoir nécessaire pour mener à bien notre étude.

En effet, nous commençons, dans ce chapitre, par présenter les outils d'imagerie médicale permettant le dépistage et le diagnostic de ce type de cancer notamment la mammographie. Ensuite, nous abordons les pathologies mammaires. Nous étudions par la suite, la classification de ces pathologies selon la norme BI-RADS (Breast Imaging-Reporting And Data System)

2.2 Imagerie mammaire

Il existe à l'heure actuelle, un certain nombre de techniques d'imagerie du corps humain couramment employées dans le domaine médical. Chacune d'elles est sensible à un type de contraste particulier et trouve ses applications pour des organes différents. Plusieurs techniques peuvent, également apporter des informations complémentaires sur un même organe. Les outils de l'imagerie médicale utilisés pour le dépistage ainsi que le diagnostic du cancer du sein sont : l'échographie (imagerie par ultrasons), l'IRM (Imagerie par Résonance Magnétique) et la mammographie (imagerie par rayons X). Dans ce qui suit, on détaille ces différentes techniques ainsi que leurs caractéristiques.

2.2.1 Échographie

Le principe de l'échographie consiste à appliquer une sonde contre la peau en regard de l'organe à explorer. Cette sonde émet des ultrasons qui traversent les tissus puis lui

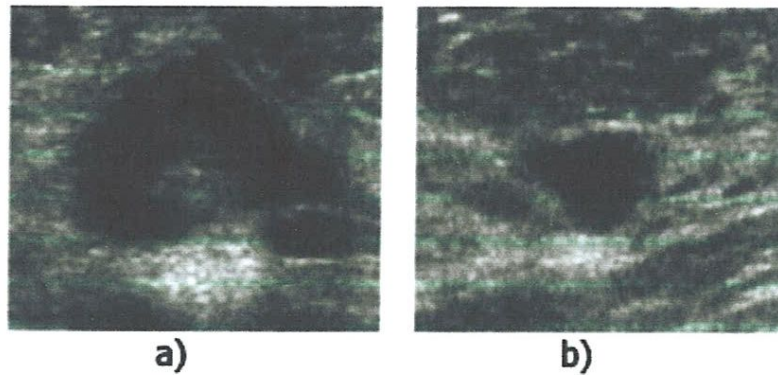


FIGURE 2.1 – *Exemples d'images échographiques du sein*[22].

sont renvoyés sous la forme d'écho. Ce signal, une fois recueilli est analysé par un système informatique qui retransmet en direct une image sur un écran vidéo. On présente dans la figure 2.1 deux exemples d'images échographiques du sein avec a) une lésion maligne et b) une lésion bénigne. La technique d'échographie présente deux avantages majeurs. Elle est d'une part peu coûteuse et d'autre part non invasive. Elle ne présente aucun risque pour la patiente, pour cette raison, elle est généralement utilisée dans le cas où la patiente est enceinte. Toutefois, vu certains inconvénients, l'échographie du sein n'est pas systématique. Elle est utilisée en complément d'une mammographie surtout qu'elle ne révèle que très rarement un cancer non détecté par la mammographie.

En outre, les micro calcifications qui sont de petites tailles(et même les petites masses) sont difficiles à détecter par ultrasons. Il est ainsi difficile de s'assurer que le sein ait été diagnostiqué dans sa totalité à l'issue de ce type d'examen. De plus , les images ultrasonores sont généralement altérées par un bruit spécifique appelé la granularité (reconnu aussi sous le nom speckle) qui est causé essentiellement par les interférences entre les ondes. En conséquence, cette méthode d'imagerie médicale n'est pas généralisée aux campagnes de dépistage. Elle est souvent exploitée comme moyen de repérage lors d'une ponction ou d'une biopsie et aussi comme moyen de repérage préopératoire pour marquer l'emplacement de la lésion.

2.2.2 Imagerie par Résonance Magnétique(IRM)

L'Imagerie par Résonance Magnétique est une technique d'imagerie médicale relativement récente (début des années 1980). Cette méthode se base sur l'utilisation d'un aimant (constituant le champ magnétique) et d'ondes de radiofréquences . Son principe consiste à faire vibrer de façon imperceptible les atomes d'hydrogène du corps humain. Placés dans un champ magnétique puissant, tous les atomes d'hydrogène s'orientent vers la même direction. Ils sont alors excités par des ondes radio durant une très courte durée. On dit

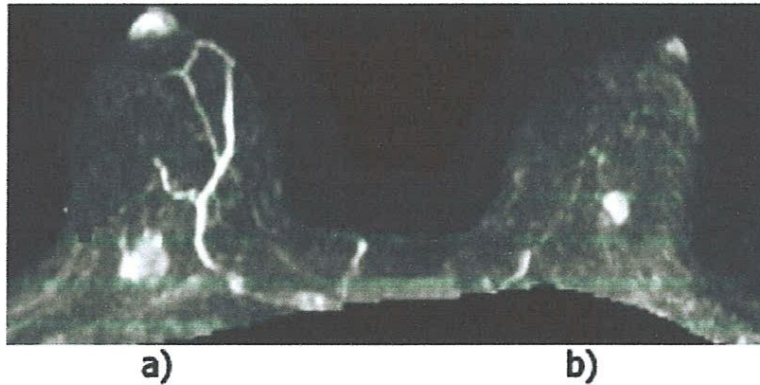


FIGURE 2.2 – *Exemple d'IRM mammaire bi-sein*[22].

qu'ils sont mis en résonance. A l'arrêt de cette stimulation, les atomes restituent l'énergie accumulée en produisant un signal. Ce dernier est enregistré et traité sous forme d'image par un système informatique. On présente dans la figure 2.2 l'image IRM des seins gauche et droit de la figure 2.1 avec a) lésion maligne et b) lésion bénigne.

L'IRM permet une étude dans tous les plans de l'espace et l'obtention d'images de grande qualité. De plus, elle fournit une image en haute résolution qui permet une étude approfondie de la pathologie. Cependant, cette technique est très coûteuse et appliquée uniquement aux patients non porteurs de prothèses métalliques. Ainsi, son utilisation est restreinte pour apporter des précisions concernant une anomalie déjà détectée en mammographie ou en échographie.

En tenant compte des conditions générales des différents moyens d'imagerie médicale (limitations de l'examen échographique, coût de l'examen d'IRM), la technique de mammographie semble être la plus adaptée à une utilisation générale dans le cadre de dépistage ou de diagnostic du cancer du sein. Nous détaillons dans ce qui suit le principe de base de ce type d'imagerie médicale.

2.2.3 Mammographie

La mammographie est une technique de radiographie, particulièrement adaptée aux seins de la femme. Elle a pour but de déceler au plus tôt des anomalies avant qu'elles ne provoquent des symptômes cliniques. Le point fort d'un tel examen est qu'il permet d'examiner la totalité du tissu mammaire avec une ou deux incidences seulement.

L'appareil dédié à la réalisation d'une mammographie est le mammographe (figure 2.3). Cet appareil se compose d'un tube radiogène générateur de rayons X de faible énergie et d'un système de compression du sein. En premier temps, les deux seins sont comprimés à tour de rôle. Cette compression permet l'étalement des tissus mammaires ce qui facilite

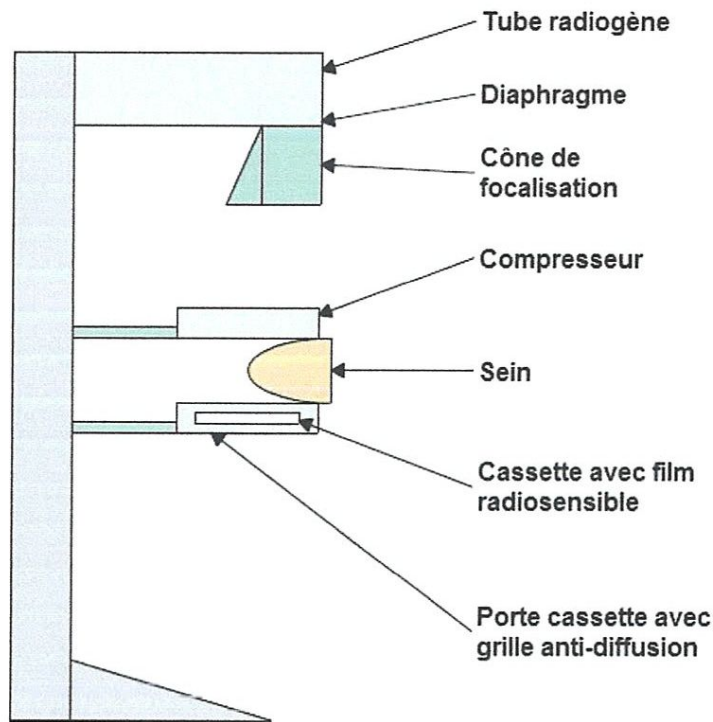


FIGURE 2.3 – Les composants d'un mammographe

la visualisation des structures du sein et la réduction de la dose de rayons X délivrée. En deuxième temps, les deux seins sont exposés à une faible dose de rayons X. On obtient, alors, une projection du sein sur un détecteur plan.

La mammographie est généralement prise sous différentes directions appelées incidences. Une bonne incidence a pour but de visualiser le maximum de tissu mammaire en l'étalant le plus possible sur la plaque radiographique. Selon la partie du sein à laquelle s'intéresse l'examen, différentes incidences sont utilisées. Les incidences les plus fréquemment utilisées sont l'incidence de face appelée aussi Cranio Caudale (CC), l'incidence oblique externe nommée Médio Latérale Oblique (MLO) et l'incidence de profil.

Dans le cas de l'incidence de face, le porte cassette (le détecteur) ainsi que le tube radiogène sont horizontaux. Cette incidence explore bien la région centrale et la région interne du sein.

Dans le cas de l'incidence oblique externe, le détecteur est incliné de 40 à 60° en fonction de la morphologie de la patiente de telle façon que l'orientation des fibres du grand pectoral soit parallèle à la cassette. L'incidence oblique médio-latérale prend en partie le muscle pectoral et permet dans ces conditions d'explorer la quasi-totalité de la glande mammaire sur une seule incidence.

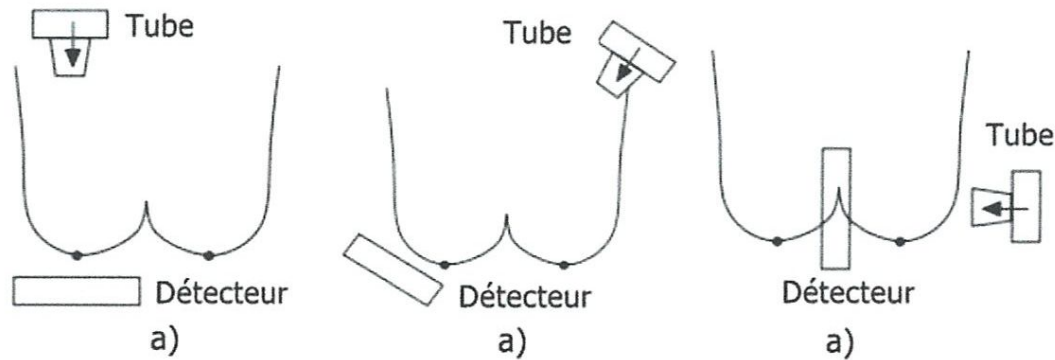


FIGURE 2.4 – Exemples d’incidences en mammographie : a) Cranio Caudale (CC), b) Médio Latérale Oblique (MLO) et c) Profil interne

Dans le cas de l’incidence de profil, le porte-cassette est vertical et placé entre les deux seins (profil interne). Cette incidence, pareillement à celle de face, explore la région centrale et interne du sein. La figure 2.4 explique le positionnement du tube radiogène et du détecteur pour les différentes incidences définies ci-dessus.

2.3 Les anomalies mammographiques

2.3.1 Les micro calcifications

Une micro calcification est un dépôt de sels de calcium composé des substances chimiques. Ces substances sont très radio-opaques et se traduisent, dans les clichés mammographiques, par de petits points clairs. Les caractéristiques qui distinguent les micro calcifications des autres éléments sont leur fort contraste et leur petite taille ($< 0,5mm$). Une fois leur taille dépasse $1mm$, on les appelle des macro calcifications et elles sont souvent bénignes. Les micro calcifications (notées Mcs) n’ont pas de taille minimale, ce qui fait que les plus petites d’entre elles peuvent facilement être confondues avec le bruit présent dans les images de mammographie.

2.3.1.1 La morphologie

L’analyse de la morphologie est très importante [27] [39] . Elle permet le plus souvent de séparer les micro calcifications bénignes et malignes. Les micro calcifications arrondies ou ovales, uniformes dans leur taille et leur forme, sont souvent bénignes. A l’inverse, celles qui sont irrégulières et hétérogènes sont souvent malignes. On détaille dans ce qui suit les différents types de micro calcifications et on donne des exemples explicatifs de chaque cas dans la figure 2.5.

- **a) Micro calcifications cutanées ou dermiques** : elles présentent typiquement un centre clair. Des clichés en incidence tangentielle sont souvent utilisés pour confirmer la localisation cutanée de ces micro calcifications.
- **b) Micro calcifications vasculaires** : ces micro calcifications en rails ou linéaires sont associées à des structures tubulaires.
- **c) Micro calcifications grossières ou coralliformes** : elles sont de grande taille (supérieures à 2 – 3mm de diamètre).
- **d) Micro calcifications en bâtonnets** : elles sont généralement associées à une ectasie canalaire (dilatation du canal galactophore) et sont alors dirigées vers le mamelon. Elles mesurent habituellement plus de 1mm de large et peuvent présenter un centre clair si le dépôt calcique se fait dans la paroi du canal.
- **e) Micro calcifications rondes** : elles ont une forme ronde et peuvent être de tailles variables. Lorsqu'elles mesurent moins de 0.5mm, elles sont dites punctiformes ou pulvérulentes.
- **f) Micro calcifications à centres clairs** : leur taille peut s'étendre de 1mm à plus de 1cm. Elles sont rondes ou ovales, à surface lisse et à centre clair. La paroi calcifiée qui les entoure est plus épaisse que celle des micro calcifications en coquille d'œuf.
- **g) Micro calcifications en coquille d'œuf ou pariétales** : ces micro calcifications très fines apparaissent comme des dépôts calciques sur la surface d'une sphère. Vu dans l'axe du rayonnement X, ce dépôt mesure généralement moins de 1mm.
- **h) Micro calcifications à type de lait calcique** : elles sont sédimentées dans le fond de kystes. En utilisant l'incidence cranio-caudale, elles sont souvent difficiles à discerner. Par contre, l'incidence de profil permet de démontrer leurs formes caractéristiques : semi-lunaires, en croissants, curvilignes ou linéaires.
- **i) Micro calcifications de suture** : elles correspondent à des dépôts calciques sur du matériel de suture. Ces micro calcifications sont typiquement linéaires ou tubulaires et présentent fréquemment des nœuds.
- **j) Micro calcifications dystrophiques** : elles mesurent habituellement plus de 0.5mm de diamètre et sont de formes irrégulières. Elles présentent parfois un centre clair. Ces micro calcifications sont souvent rencontrées dans un sein irradié ou après un traumatisme mammaire. Elles représentent la majorité des cas retrouvés en pathologie mammaire.
- **k) Micro calcifications amorphes ou indistinctes** : elles sont souvent plus ou moins rondes ou en forme de flocons. Elles sont de petites tailles et généralement à contours vagues sans forme spécifique.
- **l) Micro calcifications fines et polymorphes** : elles sont habituellement mieux visibles que les micro calcifications amorphes. Elles sont irrégulières de taille et de

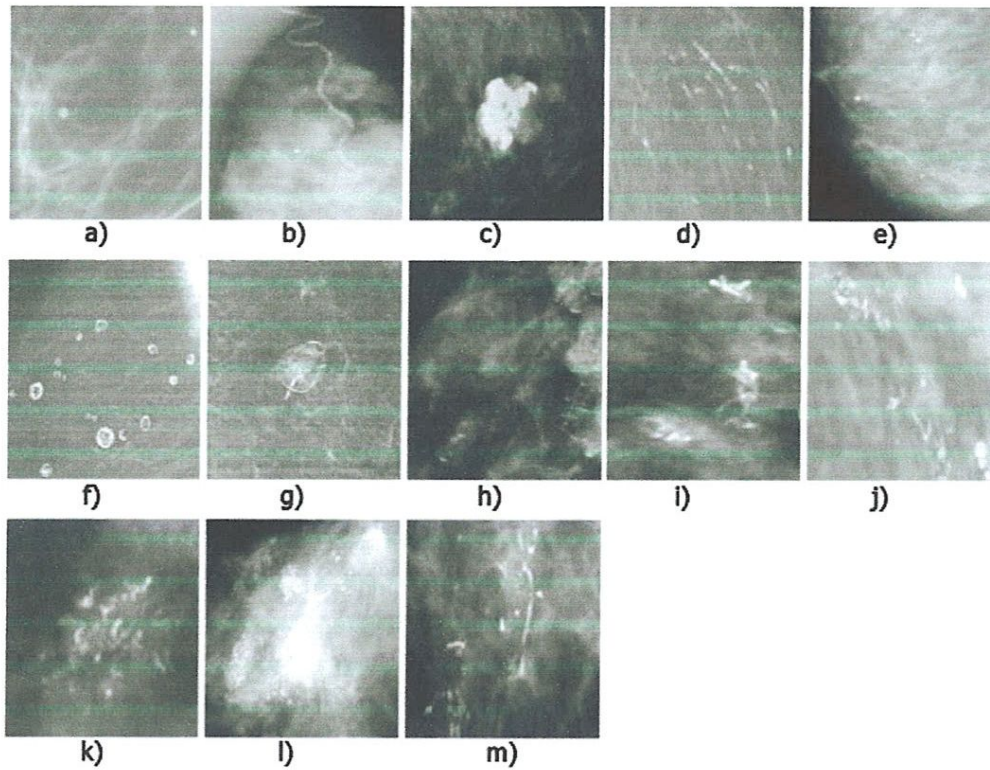


FIGURE 2.5 – *Les différents types de micro calcifications [39].*

forme variables mesurant généralement moins de $0.5mm$ de diamètre.

- **m) Micro calcifications linéaires et ramifiées** : elles mesurent moins de $0.5mm$ d'épaisseur. Elles sont irrégulières et de formes parfois linéaires ou curvilignes généralement discontinues, coudées ou branchées

2.3.1.2 Distribution

La distribution des micro calcifications est un critère fondamental. Elle présente leur répartition dans le sein et joue un rôle important dans la prise de décision de la malignité [39] . Les différentes distributions possibles des micro calcifications sont détaillées dans ce qui suit et sont représentées dans la figure 2.6.

- **a) Micro calcifications diffuses/éparses** : Dans ce cas, les micro calcifications sont distribuées de façon aléatoire dans l'ensemble du sein.
- **b) Distribution régionale** : les micro calcifications sont dispersées dans un large volume du tissu mammaire (un ou plus d'un quadrant) et ne présentent pas une distribution canalaire.
- **c) Micro calcifications groupées, en amas ou en cluster** : ces termes sont utilisés lorsque de multiples micro calcifications (au moins cinq) occupent un petit volume tissulaire.
- **d) Distribution linéaire** : les micro calcifications sont disposées les unes derrière

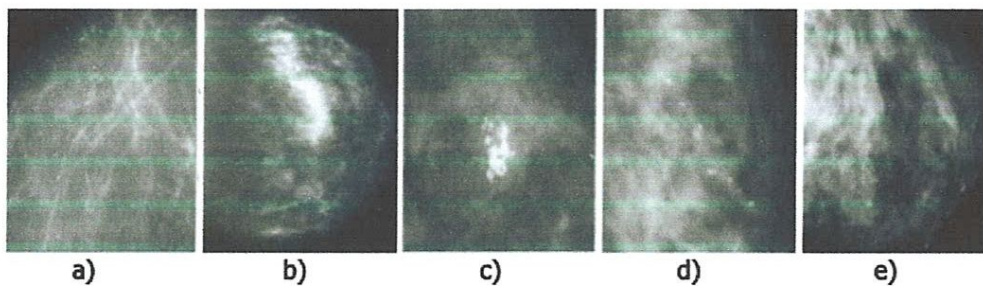


FIGURE 2.6 – *Les différentes distributions des micro calcifications[22].*

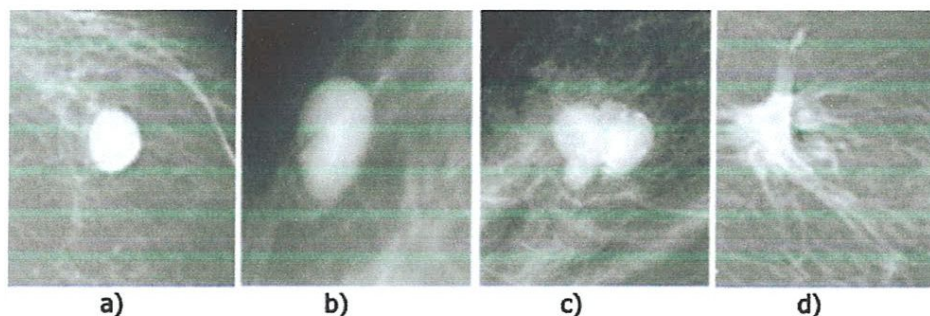


FIGURE 2.7 – *Les différentes formes possibles d'une masse[22].*

les autres sous forme d'une ligne. Il s'agit généralement de dépôts calciques dans un galactophore.

- e) **Distribution ségmentaire** : elle suggère des dépôts calciques dans des canaux galactophores ainsi que leurs branches ce qui évoque la possibilité d'un cancer mammaire étendu

2.3.2 Les masses

Une opacité ou une masse 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. Différentes caractéristiques de ces masses sont à décrire à savoir la forme, le contour et la densité .

2.3.2.1 La forme

Selon la description du BI RADS [39] , les masses mammaires peuvent avoir la forme ronde, ovale, lobulée ou irrégulière(Figure 2.7).

- a)**Ronde** : Il s'agit de masse sphérique, circulaire ou globuleuse.

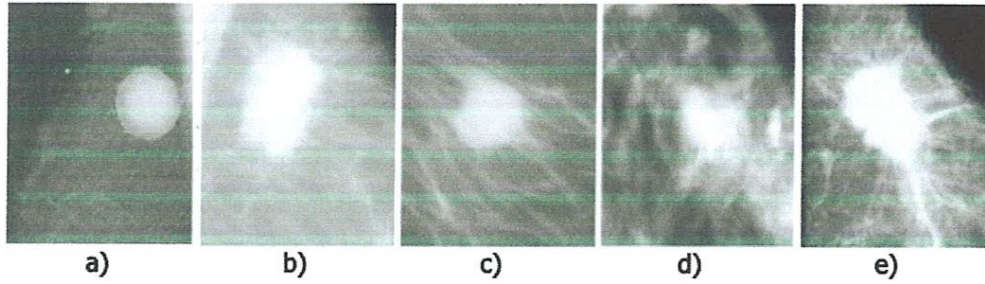


FIGURE 2.8 – Les différents contours possibles d'une masse[22].

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

2.3.2.2 Le contour

Le contour des masses mammaires est soit circonscrit, soit microlobulé, soit masqué, soit indistinct, soit spiculé. On détaille dans ce qui suit ces différentes notions (Figure 2.8) :

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

2.3.2.3 La densité

L'aspect du sein normal est très variable d'une femme à l'autre. Le facteur le plus remarquable est la grande variabilité de la densité radiologique de l'aire mammaire. Wolfe est le premier qui a établi une relation entre la densité du tissu mammaire et le risque de développer un cancer [56]. La classification BIRADS de l'ACR (American College of Radiology) définit 4 classes de la composition du sein (Figure 2.9).

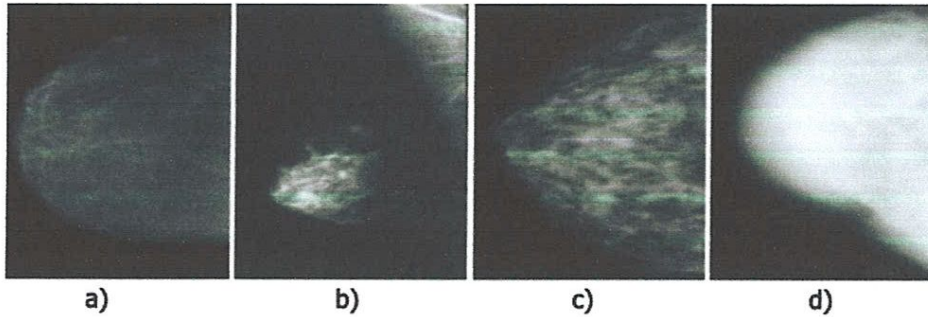


FIGURE 2.9 – *Densité mammaire selon le lexique BIRADS.*[22]

- **a) Stade 1** : Le sein est presque entièrement grasseux et homogène, radio transparent et facile à lire (moins de 25 % de la glande mammaire).
- **b) Stade 2** : Il y a des opacités fibro glandulaires dispersées. Le sein est grasseux et hétérogène (approximativement 25 à 50 % de la glande mammaire).
- **c) Stade 3** : Le tissu mammaire est dense et hétérogène (approximativement 51 à 75 % de la glande mammaire).
- **d) Stade 4** : Le tissu mammaire est extrêmement dense et homogène. La mammographie est alors difficile à interpréter puisque la densité peut masquer une lésion (plus de 75 % de la glande mammaire) (figure 2.9).

2.4 La classification des anomalies mammographiques

Il est important d'adopter un lexique standard et une classification commune afin de fournir aux radiologues une description claire et précise des lésions mammaires. Il existe plusieurs classifications comme Le Gal [18], de Lanyi [27] et de BIRADS [39].

La classification que nous avons utilisées dans notre approche est celle de BI-RADS car elle est plus récente par rapport aux autres lexiques.

2.4.1 La classification de LeGal

En 1976, Le Gal du Service de Radio diagnostique de l'institut Curie à Paris, a conçu la classification dite de Le Gal[18]. Elle décrit cinq types morphologiques qui ont une valeur prédictive de malignité croissante (Tableau 2.1) :

Cette ancienne classification a l'avantage d'être simple. Toutefois, son défaut principal est qu'elle se base uniquement sur les microcalcifications et n'intègre pas d'autres paramètres tels que :

Type 1	Mcs annulaires, arciformes ou polyédriques. Risque de cancer du sein quasi nul
Type 2	Mcs rondes et de tailles variables. Risque de carcinome : 22%
Type 3	Mcs poussiéreuses, pulvérulentes. Risque de cancer : 36%.
Type 4	Mcs irrégulières associées à un risque de cancer : 56%.
Type 5	Mcs vermiculaires ou branchées. Risque de carcinome : 90%.

TABLE 2.1 – *La classification de LeGal.*

- L'étude morphologique des masses.
- La disposition des microcalcifications.
- L'étude des distorsions architecturales.
- Le comportement du radiologue vis-à-vis de chaque cas.

2.4.2 La classification de Lanyi

Lanyi[27] individualise seulement deux types de calcifications : les calcifications rondes lobulaires ou lait calcique, associées à une pathologie bénigne et les calcifications canalaires irrégulières ou en bâtonnets en rapport avec de la nécrose et associées à un cancer.

2.4.3 La classification de BI-RADS

La classification BI-RADS est utilisée par les radiologistes lors de la mammographie, de l'échographie et de l'IRM pour définir les anomalies et permettre de savoir ce qui doit être fait par la suite, soit retour au dépistage, suivi rapproché ou biopsie. BI-RADS comporte sept catégories :

BI-RADS 0

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

BI-RADS 1

L'examen est négatif : mammographie normale. Les seins sont symétriques et il n'y a pas d'opacité, de distorsion architecturale ou de calcification suspecte. En principe cela ne pose aucun problème si les seins sont radio-transparents. Mais quelle est la certitude en

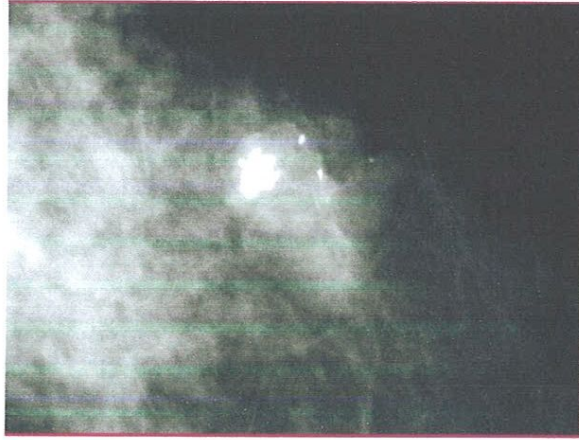


FIGURE 2.10 – *Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2 [29].*

cas de seins denses (type 3 ou 4)?.

Les solutions possibles résident dans la comparaison avec les mammographies anciennes, l'examen clinique ce qui renvoie à la question plus générale de l'intérêt de l'échographie systématiquement associée à l'exploration des seins denses.

BI-RADS 2

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

Les anomalies entrant dans cette catégorie sont les suivantes :

- Masses rondes avec macro calcifications (fibroadénome /kystes) (Figure 2.10).
- Masses rondes ou ovales à centre clair (ganglion intra mammaire).
- Masse ronde correspondant à un kyste typique en échographie.
- Anomalie de densité grasseuse ou mixte (kyste huileux, lipome, galactocèle).
- Cicatrice connue et calcification sur matériel de suture.
- Macro calcifications sans opacité (adénofibrome, kyste, calcifications vasculaires).
- Micro calcifications annulaires ou arciformes, semi-lunaires (Figure 2.11).
- Calcifications cutanées et calcifications punctiformes régulières diffuses.

BI-RADS 3

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



FIGURE 2.11 – *Microcalcifications bénignes : lait calcique : BI-RADS 2 [29].*

comporte. Il n'y a pas de modification attendue pendant la période de surveillance et les données de la littérature sont en faveur de cette attitude.

La surveillance mammographique est justifiée par le fait que ces lésions probablement bénignes ont effectivement un très faible risque d'être malignes : que la mammographie identifiera, parmi les rares lésions qui se modifient pendant l'intervalle, celles qui sont effectivement malignes après ces cancers seront diagnostiqués tôt dans leur évolution à un stade encore favorable. Ces lésions doivent être non visibles et ne pas être nouvelles ou en progression par comparaison avec une mammographie antérieure. Il faut préciser dans le compte rendu le rythme suggéré de surveillance.

Au terme d'une surveillance suffisante pendant deux à trois ans, la lésion stable peut être reclassée BI-RADS 2. En cas de modification au cours de la surveillance, la biopsie est la règle. Les données de la littérature sont en faveur de cette attitude puisque le risque de cancer observé dans la littérature est de 0,3 à 1,7 % des lésions classées BI-RADS 3. L'intérêt de l'échographie devant une masse probablement bénigne est évident : quand elle permet de faire le diagnostic immédiat de kyste ou à l'opposé d'une lésion hautement suspecte, elle évite respectivement une surveillance inutile ou une attente délétère de six mois.

Les anomalies susceptibles d'être classées BIRADS 3 sont :

- Masses ou opacités bien circonscrites rondes ou ovales ou discrètement lobulées, sans micro lobulation, non calcifiées, non typiquement liquidiennes en échographie(Figure 2.12).
- Asymétries focales de densité à limites concaves et/ou mélangées à de la graisse.
- Petits amas ronds ou ovales de calcifications amorphes, peu nombreuses, évoquant un début de calcification d'adénofibrome.
- Une discrète distorsion sans opacité centrale dans la zone d'une biopsie antérieure.

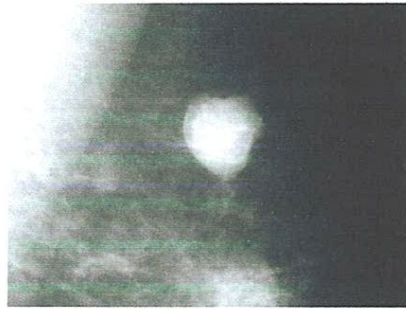


FIGURE 2.12 – *Masses bien circonscrites discrètement lobulées, sans micro lobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3 [29].*

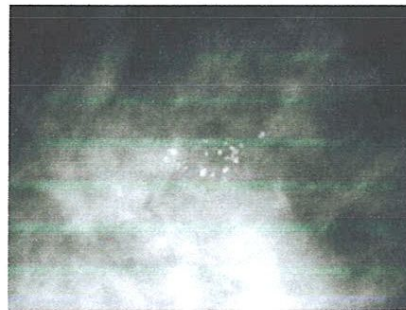


FIGURE 2.13 – *Micro calcifications granulaires irrégulières peu nombreuses : BIRADS 4 [29].*

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

BI-RADS 4

Anomalie suspecte : une biopsie doit être envisagée. Ce sont des lésions qui n'ont pas toutes les caractères morphologiques d'un cancer typique mais qui ont une probabilité importante d'être malignes [5]. Cette probabilité est de 10 à 50 % [41]. Une vérification histologique est nécessaire.

Ce sont les :

- Micro calcifications punctiformes régulières nombreuses et/ou groupées en amas aux contours ni ronds ni ovales.
- Micro calcifications amorphes, groupées et nombreuses.
- Micro calcifications irrégulières, polymorphes ou granulaires, peu nombreuses (Figure 2.13).
- Distorsion architecturale sans centre dense (Figure 2.14) en dehors d'une cicatrice connue et stable.
- Masses opaques solides (échographie) rondes ou ovales à contours lobulés ou masqués ou ayant augmenté de volume.

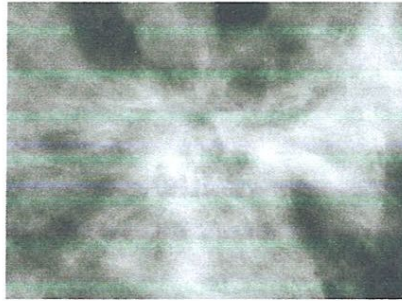


FIGURE 2.14 – Images spiculée sans centre dense : centre prolifératif d’Aschoff[29].

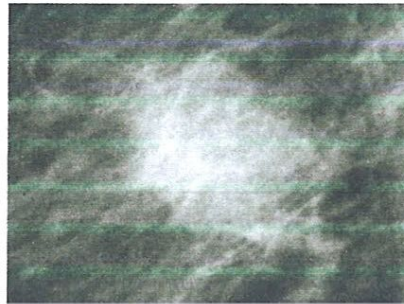


FIGURE 2.15 – Masse mal circonscrite aux contours flous et irréguliers : BI-RADS 5[29].

- Asymétries de densité focales, à limites convexes ou évolutives.

BI-RADS 5

Haute probabilité de malignité . Ces lésions ont une très forte probabilité de correspondre à un cancer du sein (> 95 %) [32].

Il s’agit de :

- Masses spiculées à centre dense.
- Masses irrégulières à contours indistincts ou spiculés (Figure 2.15).
- Micro calcifications vermiculaires, arborescentes ou micro calcifications irrégulières, polymorphes ou granulaires, nombreuses et groupées (Figure 2.16).

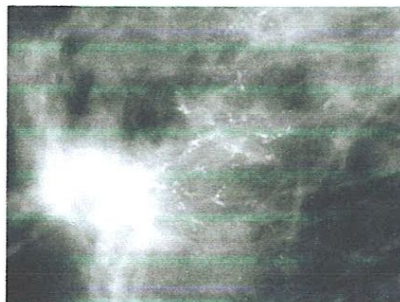


FIGURE 2.16 – Micro calcifications vermiculaires, arborescentes, nombreuses et groupées : BI-RADS 5 [29].

- groupements de micro calcifications quelle que soit leur morphologie, dont la topographie est galactophorique.
- Micro calcifications associées à une distorsion architecturale ou à une opacité.
- Micro calcifications groupées ayant augmenté en nombre ou micro calcifications dont la morphologie et la distribution sont devenues plus suspectes.

Une biopsie guidée par l'imagerie est utile pour confirmer le diagnostic, d'autant plus que le traitement envisagé comporte la technique de la chimiothérapie.

BI-RADS 6

Résultat de biopsie connu : malignité prouvée. Une action appropriée doit être entreprise.

2.4.4 Synthèse

En dépistage, pour l'évaluation statistique, toute mammographie classée dans les catégories 0, 4, 5 ou 6 est positive et toute mammographie classée dans les catégories 1, 2 ou 3 est négative. Toutefois, un diagnostic de malignité posé dans les douze mois suivant la mammographie prouve la positivité, et l'absence de cancer dans la même période prouve la négativité.

Les avantages de cette structuration du compte rendu sont pour la patiente une information intelligible, claire, loyale et pour les confrères cliniciens un compte rendu solide les aidant dans leur choix afin de diminuer le risque d'exams inutiles et de retard diagnostique.

La comparaison avec les examens antérieurs et le contexte (âge, haut risque, sein traité) peuvent faire reclasser une anomalie dans une classe inférieure ou supérieure. L'échographie va permettre de reclasser de nombreuses lésions (masses ou opacités). Lorsque la mammographie ne révèle pas la nature d'une lésion palpable, la décision d'effectuer une biopsie repose sur les données cliniques et échographiques.

2.5 Conclusion

Dans ce chapitre, les outils de l'imagerie médicale permettant le diagnostic du cancer du sein ont été présentés. Une grande attention a été consacrée à l'étude des spécifications des pathologies mammaires à savoir les masses et les micro calcifications. Une telle étude est fortement associée à la présentation des standards adoptés par les radiologues telle que BIRADS pour classer les lésions mammaires en bénignes-malignes.

Chapitre 3

Extraction des caractéristiques

3.1 Introduction

La description des lésions mammaires est une étape incontournable dans la chaîne de traitement des images mammographiques. Ces images mammographiques sont présentées en niveau de gris et ne contiennent pas d'information couleur. De ce fait, on se contente dans ce qui suit à étudier les descripteurs de texture, forme et contour.

On se focalise dans un premier temps, sur les différents descripteurs de texture et leur utilité dans le cas de traitement des masses.

Ensuite , on s'intéresse aux descripteurs de forme proposés dans la littérature. On commence d'abord par étudier les descripteurs d'ordre général qui sont utilisés dans la majorité des applications de traitement d'images. Ensuite, on aborde les descripteurs spécifiques couramment appliqués en traitement des masses mammaires. Cette étude permet de distinguer les avantages et les inconvénients des différentes méthodes de caractérisation de la forme des lésions étudiées.

A la fin, on abordons les descripteurs de contour proposés dans la littérature ainsi que les caractéristiques humaines.

3.2 Les caractéristiques des masses mammographiques

Les caractéristiques d'une masse mammographique généralement se divisent en trois grandes catégories : les descripteurs de texture, forme et contour. Dans ce qui suit un état de l'art de ces trois descripteurs.

3.2.1 Les descripteurs de texture en mammographie

Dans le domaine du traitement de l'image, il n'existe pas de définition conventionnelle de la texture. Les définitions mathématiques proposées peuvent être trop générales et donc imprécises et peuvent aussi être trop détaillées et donc ne couvrent pas tous les cas de texture envisageables. Nous citons ici quelques définitions proposées dans la littérature :

Haralick [19] a décrit la notion de texture comme suit : " Une image texturée est décrite par le nombre et les types de ses primitives tonales ainsi que leurs orientations spatiales. Elle ne peut pas être analysée sans une vue de la primitive tonale de référence. Au fur et à mesure que la résolution augmente, on observe une texture fine puis une texture grossière".

Dans sa thèse, Unser [55] donne une définition perceptuelle de la texture : " Une texture est une région d'une image pour laquelle il est possible de définir une fenêtre de dimensions minimales, telle qu'une observation à travers de celle-ci se traduit par une perception (impression) visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée ".

Il existe en littérature, plusieurs familles de textures ce qui engendre la proposition de plusieurs descripteurs de texture. En effet, certaines méthodes de description texturale basées sur une analyse fréquentielle regroupent principalement la transformée de Fourier, les filtres de Gabor et la transformée en ondelette. D'autres méthodes basées sur un modèle regroupent les champs de Markov et les mesures fractales. Une autre catégorie de méthodes basée sur une analyse spatiale et statistique des intensités de niveau de gris rassemble les statistiques du premier ordre, les caractéristiques d'auto corrélation et les matrices de co-occurrence .

Dans ce qui suit, on présente les méthodes les plus utilisées dans le domaine du traitement des images mammographies.

3.2.1.1 Les statistiques du premier ordre

Les statistiques du premier ordre mesurent la probabilité d'observer un niveau de gris à un emplacement aléatoire d'une image . Elles sont estimées sur les niveaux de gris sans la considération de leur distribution relative. Les statistiques du premier ordre sont calculées à partir de l'histogramme des intensités noté *Hist*. Cette entité est obtenue comme suit :

$$Hist(i) = \frac{1}{ng} \sum -p = 1^{ng} I(p) = i \quad (3.1)$$

avec I l'intensité des pixels de l'image, ng le nombre de niveaux de gris et $i \in \{1, 2, \dots, ng\}$. Parmi les statistiques du premier ordre, on peut citer la moyenne (Moy), l'écart type (σ_{Hist}) et la variance (Var).

— **La moyenne :**

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

— **L'écart type :**

$$\sigma_{Hist} = \sqrt{\frac{1}{ng} \sum_{i=1}^{ng} (i - Moy)^2.Hist(i)} \quad (3.3)$$

— **La variance :**

$$Var = \frac{1}{N} \sum_{p=1}^N (I(p) - Moy)^2 \quad (3.4)$$

L'inconvénient majeur des statistiques du premier ordre basées sur l'histogramme est qu'elles n'exploitent que l'information niveau de gris et ne prennent pas en considération les relations spatiales entre les pixels. La description de texture à base des statistiques d'ordre supérieur (statistiques du second ordre par exemple) s'avère plus efficace.

3.2.1.2 La matrice de co-occurrence

Les méthodes statistiques du second ordre sont les méthodes qui mettent en jeu deux pixels pour le calcul des caractéristiques. La principale méthode du second ordre utilisée pour le traitement des images médicales notamment les images mammographiques est la méthode de matrice de co-occurrence ou méthode de dépendance spatiale des niveaux de gris : SGLD (Spatial Gray Level Dependency) développée par Haralick [20]. C'est une méthode statistique qui consiste à construire des matrices de co-occurrence pour représenter les relations entre les pixels d'une image.

La matrice représente la probabilité conjointe pour deux niveaux de gris i, j soient dans une relation spatiale donnée. Cette relation est définie en termes de la distance et d'angle entre ces deux pixels. L'angle permet d'évaluer la direction de texture et l'application de plusieurs valeurs de distance peut donner une description significative de la taille de la périodicité de la texture. Le choix de la distance et de la combinaison de l'angle, ainsi que le niveau de quantification, est quelque peu arbitraire [17].

La Figure 3.1 montre la co-occurrence pour un pixel (pixel jaune) avec $d = 3$ pixels et $\theta \in \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}\}$

Dans cette méthode, il faut calculer la matrice de co-occurrence pour un ROI carré à huit angles différents : $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\}$ et à toutes les distances $d \in$

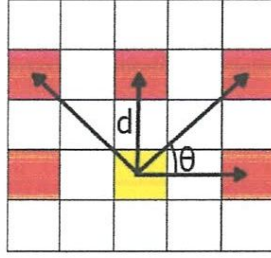


FIGURE 3.1 – La co-occurrence pour un pixel(jaune)

$\{1, 2, 3, \dots, L/2\}$, où L est la longueur de ROI. Ensuite, utiliser ces valeurs variantes de distances et des angles pour avoir des caractéristiques discriminantes pour le motif de masse de distinguer entre la masse bénigne ou maligne.

En conséquence, les huit matrices de co-occurrence sont générés pour chaque valeur de distance d , ces matrices sont additionnées pour obtenir une matrice de rotation invariant Md :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad (3.5)$$

Ensuite, extraire les huit des 14 caractéristiques de co-occurrence proposées par Haralick. Ces caractéristiques sont calculées par les équations qui suivent, où G représente le nombre de niveaux de gris, σ_i^d, σ_j^d et μ_i^d, μ_j^d désigne les écarts types et la moyenne de la somme de co-occurrence matrice à la valeur de la distance d .

Corrélation :

$$x_1 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad (3.6)$$

Homogénéité :

$$x_2 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad (3.7)$$

Variance :

$$x_3 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu^d)^2 \quad (3.8)$$

Contraste :

$$x_4 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - j)^2 \quad (3.9)$$

Énergie :

$$x_5 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad (3.10)$$

Probabilité maximale :

$$x_6 = \max_{i,j} M_{i,j}^d \quad (3.11)$$

Somme moyenne :

$$x_7 = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad (3.12)$$

Prominence de cluster :

$$x_8 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d + j - \mu_j^d)^4 \quad (3.13)$$

Appliquer ces huit mesures pour chaque valeur de $d \in [1, L/2]$, il y aura $(8 * L/2)$ caractéristiques. Ceci est très grandes quantités de données, dont la plupart sont redondantes et non utiles. Donc, il faut utiliser les fonctions suivantes pour chaque caractéristiques mesure de co-occurrences sur toutes les valeurs de d , ces fonctions sont :

Moyenne :

$$x_{n+8} = \frac{1}{L/2} \sum_{d=1}^{L/2} x_n^d \quad (3.14)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

Écart absolu moyen :

$$x_{n+16} = \frac{1}{L/2} \sum_{d=1}^{L/2} |x_n^d - x_{n+8}| \quad (3.15)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

Minimum :

$$x_{n+24} = \min_{d=1}^{L/2} x_n^d \quad (3.16)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

Maximum :

$$x_{n+32} = \max_{d=1}^{L/2} x_n^d \quad (3.17)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

Variance :

$$x_{n+40} = \frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^2 \quad (3.18)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

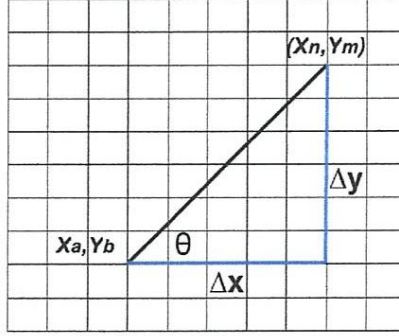


FIGURE 3.2 – Run Difference Matrix (RDM).

Skewness :

$$x_{n+48} = \frac{\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^3}{\left(\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+8})^2\right)^{3/2}} \quad (3.19)$$

pour tous $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$

Dans le cas des images mammographiques , la matrice de co-occurrence a bien été exploitée notamment pour la détection des masses et des micro calcifications et pour la classification des micro calcifications en maligne/bénigne [26]. Cependant, son emploi pour la classification des masses reste restreint.

3.2.1.3 Run Difference Matrix (RDM)

RDM est basé sur l'évaluation de la fonction de densité de probabilité des différences de niveau gris dans une image. La matrice de différence a comporté la différence niveau gris avec une distance entre les pixels, quand le vecteur de déplacement entre deux pixels est donné [34].

En ce qui concerne la figure 3.2, calculer le de vecteur de déplacement D en utilisant l'équation suivante :

$$D = [\Delta x, \Delta y] = [x_a, y_b] - [x_n, y_m] \quad (3.20)$$

RDM pourrait être défini comme un nombre total de pair de pixels dans ROI avec la distance r et les différences de niveau grise avec la direction donnée θ :

$$RDM(r, gdiff|\theta) = ((x_a, y_b), (x_n, y_m)) :$$

$$((x_a, y_b), (x_n, y_m)) \in ROI, |G(x_n, y_m) - G(x_a, y_b)| = gdiff/N \quad (3.21)$$

Où $G(x, y)$ est la valeur de niveau grise du pixel (x, y) :

$$r = \sqrt{\Delta x^2 + \Delta y^2} \quad (3.22)$$

$$\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right) \quad (3.23)$$

et N est utilisé pour la normalisation et c'est le nombre total de toutes les paires de pixels dans ROI :

$$N = ((x_a, y_b), (x_n, y_m)) : ((x_a, y_b), (x_n, y_m)) \in ROI \quad (3.24)$$

Dans des cas de cancer, la forme et les marges sont non uniformes. Donc il faut prolonger les matrices RDM pour huit angles différents $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\}$. En conséquence, huit matrices RDM sont produites, ces matrices sont additionnés pour obtenir une rotation la matrice invariable M_{RDM} :

$$M_{RDM} = \sum_{\theta} RDM(r, gdiff|\theta) \quad (3.25)$$

La matrice obtenue M_{RDM} n'est pas conçue pour l'extraction de caractéristiques mais il est possible d'utiliser les trois vecteurs caractéristiques définis avec l'original la matrice de différence exécutée pour extraire des caractéristiques.

Ces caractéristiques sont le (DGD) (la distribution de différence de niveau grise), le vecteur de (DOD) (la distribution de différence moyenne) et le vecteur de (DAD) (la distribution de distance moyenne) :

$$DGD_{gdiff} = \sum_{r=1}^{L/2} M_{RDM} \quad (3.26)$$

$$DOD = \sum_{gdiff}^{G-1} M_{RDM.gdiff} \quad (3.27)$$

$$DAD_{gdiff} = \sum_{r=1}^{L/2} M_{RDM.r} \quad (3.28)$$

De ces trois caractéristiques, il est possible d'extraire autres caractéristiques tel que :

a) Large difference emphasis (LDE) : qui mesure la prédominance des grandes différences de niveaux de gris.

$$x_{57} = LDE = \sum_{gdiff=0}^{G-1} DGD(gdiff) \cdot \ln(2/gdiff) \quad (3.29)$$

b) **Sharpness** : qui mesure le contraste et la définition d'une image.

$$x_{58} = Sharpness = \sum_{gdiff=0}^{G-1} DGD(gdiff).(gdiff)^3 \quad (3.30)$$

c) **Second Moment de DGD (SMG)** : qui mesure la variation des différences de niveau de gris.

$$x_{59} = SMG = \sum_{gdiff=0}^{G-1} (DGD(gdiff))^2 \quad (3.31)$$

d) **Long distance emphasis for large difference (LDEL)** : qui mesure la proéminence des différences large sur une longue distance.

$$x_{61} = LDEL = \sum_{gdiff=0}^{G-1} DAD(gdiff)(gdiff)^2 \quad (3.32)$$

3.2.1.4 Gray Level Co-occurrence Matrix (GLCM)

Dans l'analyse statistique de la texture, les caractéristiques de texture sont calculées à partir de la distribution statistique des combinaisons d'intensités observées à des positions spécifiées par rapport à l'autre dans l'image. Selon le nombre de points d'intensité (pixels) dans chaque combinaison, les statistiques sont classées en premier ordre, de second ordre et d'ordre supérieur.

La méthode Gray Level Co-occurrence Matrix (GLCM) est une manière d'extraire les caractéristiques statistiques de texture du second degré .Un GLCM est une matrice dont le nombre de lignes et de colonnes est égal au nombre de niveaux de gris g dans l'image. L'élément de la matrice $P(i, j|j, \theta)$ est la fréquence relative avec laquelle deux pixels séparé par la distance d et direction indiquée par l'angle particulier (θ) , une avec une intensité i et l'autre avec une intensité j .

L'algorithme de GLCM de base est le suivant [46] :

- Compter toutes les paires de pixels dans lequel le premier pixel a une valeur i et sa paire de correspondance déplacée du premier pixel par D a une valeur de j .
- Ce nombre est inscrit dans la colonne i ème rangée et j du Pd matrice $[i, j]$.
- On notera que $Pd [i, j]$ ne est pas symétrique, car le nombre de paires de pixels ayant des niveaux de gris $[i, j]$ ne est pas nécessairement égal au nombre de paires de pixels ayant des niveaux de gris de $[j, i]$.
- Les éléments de $Pd [i, j]$ peuvent être normalisés en divisant chaque entrée par le nombre total de paires de pixels.

— normalisé GLCM $N [i, j]$, définie par :

$$N[i, j] = \frac{P[i; j]}{\sum_i \sum_j P[i; j]} \quad (3.33)$$

3.2.1.5 La transformée de Fourier

La transformée de Fourier est l'une des méthodes utilisées pour l'analyse de texture d'une image. A partir de la transformée de Fourier, on calcule généralement le spectre de puissance qui est défini comme le carré du module de la transformée de Fourier et qui est, d'après [43] très informatif sur la périodicité d'une texture.

La transformée de Fourier est un outil bien adapté à l'analyse des comportements harmoniques. Cependant elle n'est pas appropriée à l'analyse de formes localisées spatialement. Par ailleurs, la transformée de Fourier est difficilement assimilable par les chercheurs étant donné qu'il est difficile d'exploiter et d'interpréter directement ses valeurs.

3.2.1.6 L'analyse fractale

La mesure fractale est reconnue pour sa grande capacité à caractériser la rugosité. La courbe de Koch représente l'une des premières courbes fractales à avoir été décrite. Elle a été inventée en 1906 par le mathématicien suédois Helge von Koch. Parmi les méthodes d'analyse fractale, la méthode la plus populaire est celle de comptage de boîtes (box counting).

La dimension fractale d'un ensemble A est définie comme suit :

$$D = \log(N) / \log\left(\frac{1}{r}\right) \quad (3.34)$$

Telle que N est le nombre total des copies distinctes similaires à A et $\frac{1}{r}$ correspond au facteur d'échelle avec lequel A est divisé.

Une deuxième méthode basée sur le mouvement Brownien fractionnaire (fractional Brownian motion) est souvent utilisée pour décrire des phénomènes aléatoires. L'avantage de la dimension fractale du mouvement Brownien fractionnaire est son invariance aux transformations linéaires telles que la translation.

L'analyse fractale est généralement utilisée pour générer des textures ainsi que pour les analyser. Cependant, leur utilisation pour la caractérisation texturale est une tâche assez difficile puisqu'elle nécessite de définir de nombreux paramètres souvent complexes à manipuler.

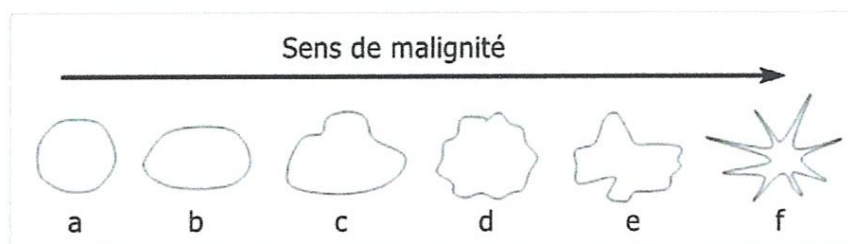


FIGURE 3.3 – Relation entre complexité du contour et malignité : a) masse circonscrite circulaire, b) circonscrite ovale, c) macro lobulée, d) micro lobulée, e) irrégulière, f) spéculée.

De plus, de telles méthodes sont souvent coûteuses en termes de temps de calcul. Dans le domaine du traitement des images mammographies, particulièrement le cas de classification de lésions de type masse, cette mesure est appliquée sur le contour en tant que descripteur de forme et non en tant que descripteur de texture qui analyse l'intérieur de la forme.

Les descripteurs de texture sont assez variés et sont couramment utilisés dans le domaine de la mammographie. Les descripteurs de texture ont prouvé leur efficacité dans le traitement de ces petites lésions qui se distinguent par leur forme et par leur répartition dans le sein.

Même si une différence de texture a été révélée entre les masses malignes et celle bénignes, d'après la littérature cette différence n'est pas assez importante et consistante pour en tenir compte [46] [49]. Par ailleurs, l'étude des caractéristiques pathologiques des masses mammaires montre que la malignité est fortement liée à la forme et au contour de ces lésions. Par conséquent, on s'intéresse, désormais, aux descripteurs de forme appliqués aux masses mammaires.

3.2.2 Les descripteurs de forme en mammographie

Le degré de malignité dans le cas des masses augmente essentiellement avec l'élévation de la complexité du contour. Pour cette raison, il serait plus judicieux d'étudier chaque cas séparément.

Nous résumons dans la figure 3.3, les différentes caractéristiques de forme et de contour d'une masse nous permettant de mieux comprendre la nature des descripteurs de forme les plus adéquats à chercher. La forme des masses mammaires peut être modélisée à travers le contour ou bien l'intérieur de l'objet. Les deux approches sont pratiquement équivalentes étant donné que la forme peut être reconstituée à partir de son intérieur et que l'intérieur peut être reconstitué à partir de son contour.

Quel que soit la façon par laquelle la forme a été modélisée, les descripteurs de formes proposés dans la littérature se basent sur des caractéristiques de la forme elle-même. Après plusieurs décennies de recherche dans le traitement d'images, une multitude de descripteurs de forme sont proposés. Nous détaillons dans ce qui suit ces différents types de descripteurs et leurs intérêts.

3.2.2.1 Les descripteurs géométriques

Les descripteurs géométriques permettent de caractériser l'aspect de la forme d'une façon plus ou moins globale et sont généralement indépendants du domaine d'application comme l'aire et le périmètre

a) Le centre de gravité : Cette mesure est étroitement liée a la forme de l'objet, de telle sorte que les coordonnées (x_g, y_g) du centre de gravité sont définies de la manière suivante :

— si la forme est représentée par toute la région de l'objet :

$$x_g = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.35)$$

$$y_g = \frac{1}{N} \sum_{i=1}^N y_i \quad (3.36)$$

avec N le nombre de points dans la région de la forme et (x_i, y_i) sont les points contenus dans cette région.

— si la forme est représentée par le contour de l'objet :

$$x_g = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad (3.37)$$

$$y_g = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad (3.38)$$

avec N le nombre de points dans le contour de la forme et (x_i, y_i) sont les points contenus dans ce contour.

b) L'aire : c'est le paramètre de forme le plus trivial qui peut être calculé à partir d'un objet détecté sur une image. Il peut être définie comme étant le nombre de pixels contenus dans une liésion. On notera ce paramètre par A [7]

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

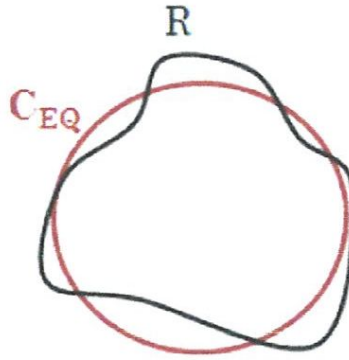


FIGURE 3.4 – Calcul de la mesure de circularité

contour.

d) La circularité : La mesure de circularité représente le degré de ressemblance entre la forme considérée et un cercle. Ce fameux descripteur noté C peut être calculé en dessinant un cercle centré sur le centre de gravité de la région et de même aire.

Soit R : la région de masse et C_{EQ} : le cercle équivalent de même aire que la zone de masse et ayant comme centre le centre de gravité de R .

La mesure de circularité peut alors se définir comme suit [60][10][50][58] :

$$C = \frac{\text{aire}(R \cap C_{QE})}{\text{aire}(R)} \quad (3.39)$$

La figure 3.4 explique d'avantage le principe de calcul de la circularité. Cette mesure est d'une valeur comprise entre 0 et 1. La région est d'autant plus circulaire que sa mesure de circularité est proche de 1.

En dépit de sa simplicité, ce descripteur est très utilisé dans plusieurs domaines d'analyse de forme. Cependant, son utilité dans le diagnostic du cancer du sein réside dans la nature du caractère pathologique des masses mammaires.

e) La rectangularité : Un autre descripteur de forme intitulé rectangularité ou boîte englobante est souvent utilisé pour décrire le taux de rectangularité et d'allongement d'une région [10][58][60].

On désigne par R la région étudiée et la boîte englobante (BE) qui est définie par le plus petit rectangle contenant la lésion étudiée (figure 3.5). La mesure de rectangularité

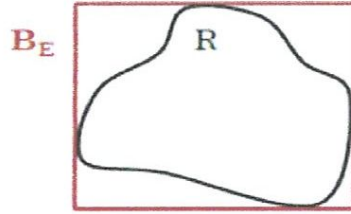


FIGURE 3.5 – Calcul de la mesure de rectangularité.

notée $Rect$ peut alors se définir comme suit :

$$Rect = \frac{aire(R)}{aire(B_E)} \quad (3.40)$$

Selon cette formule, plus l'objet à décrire est rectangulaire, plus la valeur de rectangularité est proche de 1. Toutefois, cette formule de rectangularité ne tient pas compte de l'orientation de l'objet. En effet, la valeur de rectangularité d'un rectangle parfait et droit est égale à 1 alors que la rectangularité de ce même rectangle mais considéré incliné est différente de 1. Ainsi, ce descripteur est sensible à la rotation.

f) La compacité : Il s'agit d'une mesure de la complexité du contour vis à vis de l'aire [44]. Elle est notée par Com et est donnée par :

$$Com = \frac{P^2}{A} \quad (3.41)$$

Où P est le périmètre de l'objet et A est l'aire en pixels.

Cette valeur de compacité permet de distinguer une forme irrégulière d'une forme simple puisqu'elle attribue à la forme irrégulière une valeur plus élevée. En dépit de sa simplicité, ce descripteur présente l'avantage d'être invariant aux transformations géométriques telles que la translation, la rotation et la variation d'échelle. Pour cette raison, ce descripteur servira plus tard de comparateur pour évaluer les nouvelles propositions de descripteurs.

g) La méthode de coque convexe : Au cours des dernières années, les chercheurs ont utilisé de nombreux types des caractéristiques pour distinguer entre les masses malignes et bénigne[51]. Dans cette étude [3], six nouvelles caractéristiques de forme étaient extrait, ceux-ci sont définies comme suit :

1. Zone de différence : zone de coque convexe moins la zone de la masse .

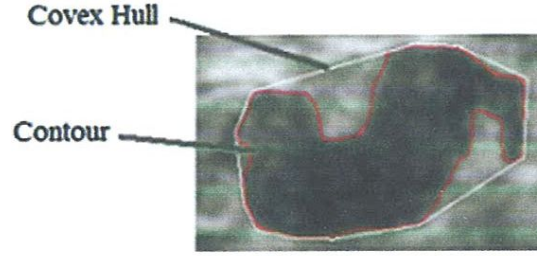


FIGURE 3.6 – La coque convexe et le contour d’une masse mammaire[3].

Pour définir autres caractéristiques, une fonction nommée ” variation ” est tout d’abord défini comme suit : Variation : la projection du distance entre les pixels les plus éloignés d’une région de masse à tous les angles.

Cinq caractéristiques sont extraites de la fonction de variation comme :

2. Moyen de variations : la valeur moyenne de la fonction de variation.

$$M_V = \frac{1}{L} \sum_{i=1}^L \text{variation}(i) \quad (3.42)$$

3. Variance de variations : valeur de la variance de la fonction de variation.

$$VAR_V = \frac{1}{L} \sum_{i=1}^L (\text{variation}(i) - x6)^2 \quad (3.43)$$

4. Skewness de variations : valeur d’asymétrie de la fonction de variation.

$$SK_V = \frac{\frac{1}{L} \sum_{i=1}^L (\text{variation}(i) - x6)^3}{\left(\frac{1}{L} \sum_{i=1}^L (\text{variation}(i) - x6)^2\right)^{3/2}} \quad (3.44)$$

5. Kurtosis de variations : valeur de l’aplatissement de la fonction de variation.

$$K_V = \frac{(1/L) \sum_{i=1}^L (\text{variation}(i) - x6)^4}{\left((1/L) \sum_{i=1}^L (\text{variation}(i) - x6)^2\right)^2} - 3 \quad (3.45)$$

6. Entropie de variations : valeur entropie de la fonction de variation.

$$EN_V = - \sum_{i=1}^L \text{variation}(i) \log_2 \text{variation}(i) \quad (3.46)$$

La coque convexe , ce qui est le plus petit polygone convexe qui contient le contour de la masse et de la région de masse, est représenté sur la Fig 3.6. La zone de la coque convexe est le nombre réel de pixels dans la coque convexe de la masse , et la zone de

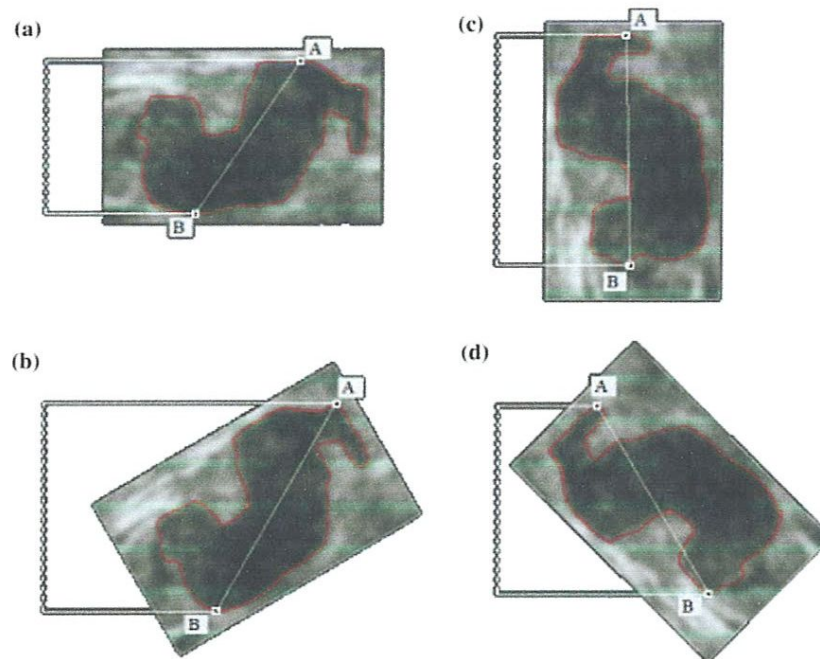


FIGURE 3.7 – *Un exemple de caractéristiques 2-6 à l'aide d'une masse en rotation à des angles différents[3].*

masse est le nombre de pixels dans la région de masse.

La figure 3.7 illustre la base de caractéristiques 2-6 . Pour ces caractéristiques, la sous-image ROI devrait être mis en rotation à différents angles. Rotation avec 5° intervalles est considéré. Pour chaque angle des pixels avec des coordonnées minimales Y ('a' dans la figure) et les coordonnées maximales Y ('b' sur la figure) sont trouvé. La projection de la distance entre les pixels sur l'axe Y est calculé et affiché avec une ligne en pointillés dans la figure.

3.2.2.2 Les descripteurs spécifiques

Les chercheurs dans le domaine de la reconnaissance de forme se sont penchés sur la proposition de descripteurs dits spécifiques, plus précis et plus informatifs sur les détails de la forme. Ces propositions visent l'obtention de descripteurs robustes capables de révéler l'aspect général d'un objet tout en préservant ses caractéristiques partielles.

a) La courbure : La courbure noté $Curv$ a été couramment utilisée dans le contexte de l'analyse de formes dans plusieurs domaines. Elle a été reconnue pour sa capacité à caractériser les formes d'objets. De nombreuses approches proposées suggèrent minimisant la courbure tout en respectant les contraintes géométriques de directions tangentes calculé sur le contour. Cette méthode est robuste et utilisée dans le système de (CAD). En général, la courbure à un point donné A , d'une courbe est définie comme l'inverse du rayon du

cercle osculateur dans A . Le cercle osculateur peuvent être obtenues comme suit : étant donné que deux points B et C à proximité de A , nous calculons le seul cercle par A , B et C . Si ces points sont alignés, le cercle a un rayon infini et la courbure est alors nulle.

$$Curv = \frac{1}{R} \quad (3.47)$$

le rayon du cercle osculateur est défini comme suit :

$$R = \frac{a.b.c}{\sqrt{(a+b+c)(a-b+c)(a+b-c)(b-a+c)}} \quad (3.48)$$

avec $a = |AB|$, $b = |BC|$ et $c = |AC|$

Récemment, plusieurs méthodes de calcul de la courbure ont été proposées. Nguyen et al[37] ont amélioré l'estimation des cercles osculateurs proposée par [12] en utilisant des segments flous. Ainsi, l'estimation proposée est plus adaptée aux contours bruités et plus significative dans le cas de points non connexes. Une autre approche proposée par Kerautret and Lachaud [24] suggère de minimiser la courbure en respectant les contraintes géométriques issues des directions tangentes calculées sur le contour

b) La longueur radiale normalisée (LRN) : Kilday et al[25] a élaboré un ensemble de six caractéristiques de forme basé sur la longueur Radial normalisée (LRN). La longueur radiale est définie comme la distance euclidienne entre le centre de gravité de l'objet et l' i^{me} pixel du contour. Ainsi, la longueur radiale normalisée est le résultat de normalisation par le maximum de distance trouvée.

$$d(i) = \frac{\sqrt{(x(i) - x_g)^2 + (y(i) - y_g)^2}}{\max(d(i))}, i = \{1, 2, \dots, N\} \quad (3.49)$$

avec $(x(i), y(i))$ et (x_g, y_g) les coordonnées de i^{me} pixel et le centre de gravité respectivement. N le périmètre de masse.

1. La moyenne de LRN (d_{avg}) : La moyenne de (LRN) est définie comme suit :

$$d_{avg} = \frac{1}{N} \sum_{i=1}^N d(i) \quad (3.50)$$

2. La déviation standard de LRN(σ) : Cette mesure décrit l'irrégularité :

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (d(i) - d_{avg})^2} \quad (3.51)$$

3. l'Entropie (E) : L'entropie est obtenu à partir de l'histogramme de la longueur radiale. Le périmètre p_k est la probabilité que (LRN) est entre $d(i)$ et $d(i) + \frac{1}{N_{bins}}$, avec N_{bins} numéro de l'histogramme normalisée, variant dans l'intervalle $[0, 1]$ qui a été divisé en $N_{bins} = 100$. La mesure de l'Entropie intègre simultanément la notion de circularité et irrégularité :

$$E = \sum_{k=1}^{100} p_k \log(p_k) \quad (3.52)$$

4. Le rapport de surface (A_1) : Le rapport de surface est une mesure du pourcentage de la partie de l'objet hors la région circulaire définie par la moyenne de(LRN) :

$$A_1 = \frac{1}{d_{avg} \cdot N} \sum_{i=1}^N (d(i) - d_{avg}) \quad (3.53)$$

5. La rugosité(R) : La rugosité vise à isoler la forme macroscopique de l'objet a partir de la structure fine des bords. Il fournissent des informations sur la moyenne entre les pixels voisins, les contours irréguliers fournissent des valeurs élevées de rugosité.

$$R = \frac{1}{N} \sum_{i=1}^N (d(i) - d(i+1)) \quad (3.54)$$

6. Le taux de croisement en zéro(ZC_1) : Il calcule le nombre de fois que la droite définie par la moyenne de (LRN) intercepte la lésion de contour, il fournit le degré de spéculation du contour.

c) La longueur Radial normalisée modifiée(LRNM) : À partir des propriétés de (LRN) Chen et al[11] ont proposés des descripteurs améliorées qui ont montré des performances supérieures à celles de base fonctions(LRN). Les nouveaux descripteurs extraits à partir de la (LRNM) sont la différence entre déviation standard (σ_{diff}), l'entropie de

la différence entre $d(i)$ et $d_{ma}(i)$ nommé E_{diff} , la zone ratio A_2 et le nombre de passage à zéro (ZC_2)

1. Différence des déviations standard (σ_{diff}) : σ_{diff} peut estimer le degré d'irrégularité du contour, de sorte que, si le contour devient de plus en plus irrégulier, σ_{diff} atteint les plus hautes valeurs.

$$\sigma_{diff} = |\sigma - \sigma_{ma}| \quad (3.55)$$

2. La différence de l'entropie (E_{diff}) : Ce descripteur fournit la distribution de la différence entre $d(i)$ et $d_{ma}(i)$. p_k est la probabilité que $|d(i) - d_{ma}(i)|$ est entre $|d(i) - d_{ma}(i)|$ et $|d(i) - d_{ma}(i)| + 1/N_{bins}$.

$$E_{diff} = \sum_{k=1}^{100} p_k \log(p_k) \quad (3.56)$$

3. Le rapport de surface modifiée (A_2) : La description est obtenue par l'équation suivante :

$$A_2 = \frac{1}{d_{avg} \cdot N} \sum_{i=1}^N (d(i) - d_{ma}(i)) \quad (3.57)$$

4. Le taux de croisement en zéro modifié (ZC_2) : C'est la mesure du nombre de fois que la courbe $d(i)$ intercepte $d_{ma}(i)$.

Finalement on présente dans le tableau 3.1 les avantages et les inconvénients de chaque descripteurs géométriques et spécifiques définis dans les sections auparavant.

3.2.3 Les descripteurs de contour en mammographie

Les systèmes d'aide au diagnostic sont largement utilisé pour la détection de différentes types d'anomalies dans les images mammographiques. Les masses sont l'une de ces types d'anomalies qui sont principalement caractérisés par leur marge et forme.

Toutefois, le nombre des caractéristiques qui décrire le contour est beaucoup moins que les caractéristiques de la forme et la texture. Dans Bagheri et all [2] ont présenté des nouveaux caractéristiques pour décrire le contour de la masse. Lorsque la limite segmentée est disponible, l'analyse de marge peut être effectuée. Tout d'abord, un ensemble d'ondelites doivent être extraites, que soit la marge de masse. L'idée de lignes radiales (pas exactement les ondelettes présentées ici) ont également été présentés dans [13] mais pour différentes raisons. Soit $\delta\Omega$ la marge de masse, le chemin de la marge est traversé et les ondelltes sont placés chaque 5 pixels (marge est sous-échantillonnée au taux 5). Q_1 sont les coordonnées du centre de gravité de la forme et est calculé comme :

$$Q_1 = \frac{1}{|\Omega|} \sum_{i,j \in \Omega} (i, j) \quad (3.58)$$

<i>Les descripteurs</i>	<i>Les avantages</i>	<i>Les inconvénients</i>
<i>A et P</i>	-Plus performants quand ils sont associés à d'autres descripteurs	-Insuffisants pour décrire les masses complexes -Non invariants par homothétie -Non uniques
<i>Com</i>	-Proportionnel à la complexité de la forme -Invariant	-Non unique
<i>C</i>	-Simple -Distingue les masses circulaires -Invariant	-Insuffisant pour décrire des masses complexes
<i>Rect</i>	-Distingue les masses ovales	-Non invariant par rotation -Insuffisant pour décrire des masses complexes
<i>LRN et LRNM</i>	-Invariants -Proportionnels à la complexité de la forme	-Dépendent de la distance entre le contour et le centre de gravité -Sensibles au bruit -Non unique
<i>Curv</i>	-Proportionnel à la complexité de la forme	-Non Invariant (homothétie) -Sensibles au bruit

TABLE 3.1 – *Les avantages et les inconvénients des descripteurs géométriques et spécifiques*

Où Ω indique Région de la forme, Si Q_2 est le pixel qui est sélectionnée en traversant la marge, puis l'ondelette l auront le même sens que la ligne $\overline{Q_1Q_2}$. La longueur de l est calculé par :

$$|l| = 2^{(\lfloor \log_2 r \rfloor) + 1} \quad (3.59)$$

Où r est la longueur de $\overline{Q_1Q_2}$. Ainsi, la longueur de l est une puissance de deux et approprié pour la transformation en ondelettes, jusqu'à présent, la longueur et le sens de l ont été déterminés, Il devrait maintenant être localisé de telle sorte que son élément central est Q_2 . La figure 3.8 montre un exemple de deux ondelettes possibles l_1 et l_2 , où Q_2^1 et Q_2^2 sont leurs pixels qui correspondent sur le contour, un exemple réel de la construction des ondelettes est illustré dans la figure 3.9

Les ondelltes sont la base pour l'analyse de la marge des masses dans la méthode proposé, a partir de ce point, il faut concentrer sur l'analyse de ces signaux pour obtenir des informations sur marge. A cette effet, la transformation en ondelettes unidimensionnel a été appliqué à chacune d'entre elles utilisant le filtre de haar. Soit n le nombre de ondelettes et l_i le ième ondelette. Après L niveau de décomposition, les coefficients d'ondelettes

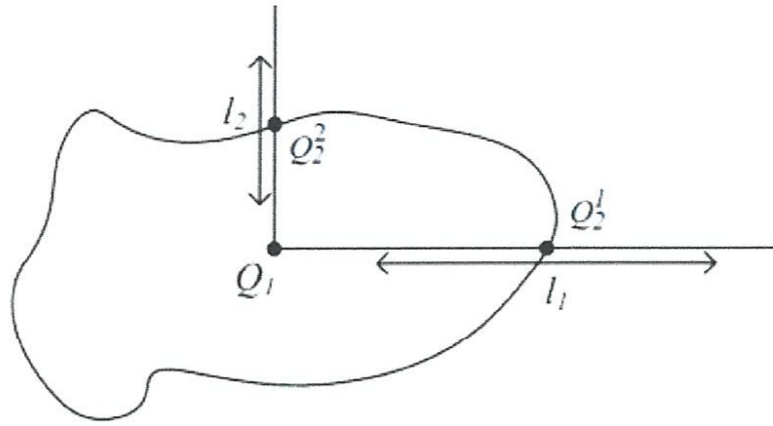
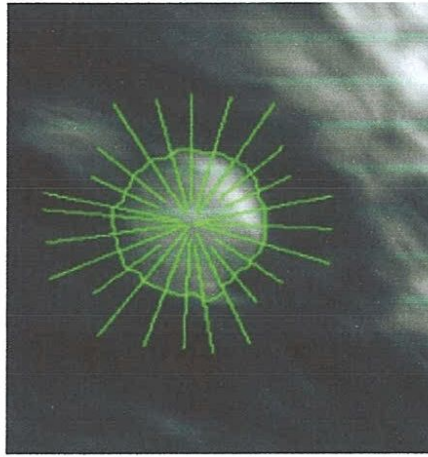

 FIGURE 3.8 – Exemple de deux ondelettes l_1 et l_2 .


FIGURE 3.9 – Capture des ondelettes avec le taux 10 sur le contour[2]

$\{a_L, d_L, d_{L-1}, \dots, d_1\}$ sont obtenue. L est calculé comme :

$$L = \log_2 |l_i| - 2 \quad (3.60)$$

où $|l_i|$ est la longueur d'onde de la i -ième. Il faut seulement utilisé les coefficients de détail et en raison de la texture ambiguë de mammographies d_1 a été abandonné. Les valeurs de chaque niveau sont normalisées séparément pour être entre 0 et 1.

A l'aide de coefficients d'ondelettes, la probabilité de distribution de contour est calculée pour chaque l_i . Supposons que D est la longueur de d_2 , puis la fonction de probabilité de masse (PMF) de bords a également D éléments. La probabilité de bords P_i^j (j ième élément de P_i) est calculée comme :

$$P_i^j = S_j / \sum_{j=1}^D S_j \quad (3.61)$$

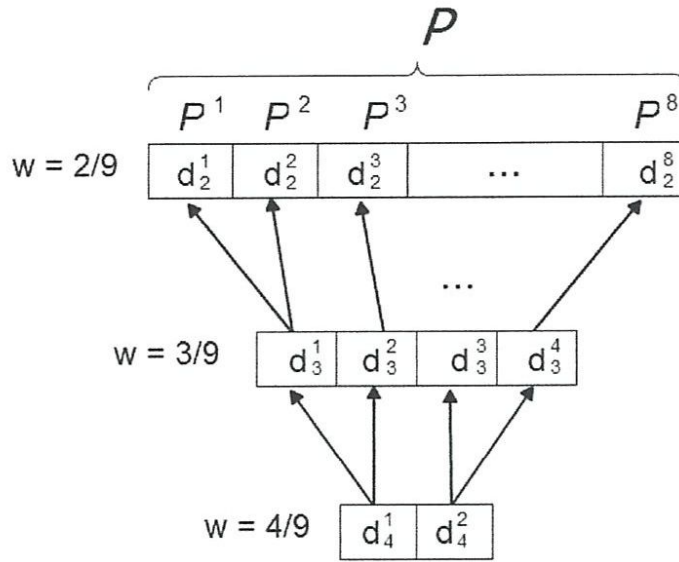


FIGURE 3.10 – *Modèle schématique du calcul des probabilités de contour de chaque forme d'onde*

$$S_j = \sum_{t=2}^L w_t d_t^{u(t,j)} \quad (3.62)$$

$$u(t, j) = \left\lfloor \frac{j}{2^{t-2}} \right\rfloor \quad (3.63)$$

où d_t^u est le u ème élément du t ème niveau de détails dans la décomposition, w_t est le poids qui est associé à la t ème niveau de détails et calculé comme :

$$w_t = t / \sum_{i=2}^L t \quad (3.64)$$

Le plus faible poids assigné au deuxième niveau et le poids le plus élevé assigné à l' L ème niveau de décomposition, et $\sum_t w_t = 1$. La raison du choix de ces poids, c'est que l'énergie dans des niveaux supérieurs est bien plus que dans des niveaux inférieurs dans la décomposition de chaque forme d'ondelettes. En outre, l'effet de bruit est aussi plus sensible aux niveaux inférieurs. Pour plus de clarification de l'équation (3.61-3.63) un exemple schématique est illustré dans la figure 3.10 où $L = 4$ et $D = 8$ chaque P^j est calculé par la multiplication de l'élément correspondant par ses poids et après l'addition au sens des flèches. Enfin, pour la contrainte de $\sum_j P^j = 1$ les éléments de P sont normalisés à leurs sommes qui est déclarée dans l'équation (3.61).

Avant de mettre fin à la discussion concernant le calcul (PMF), il vaut de mentionner que la distribution de probabilité discrète a été calculé. On peut aller plus loin et calculer une distribution en utilisant des techniques comme Parzen windows et des k-Nearest Neighbors(KNN), Cette simple distribution discrète est adéquate.

Pour chaque fonction de distribution de P_i qui a été obtenu à partir de l_i les paramètres suivants sont calculés :

3.2.3.1 Kurtosis :

$$k_i = \frac{(1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^4}{((1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^2)^2} - 3 \quad (3.65)$$

3.2.3.2 Entropie :

$$E_i = - \sum_{j=1}^D P_i^j \log_2 P_i^j \quad (3.66)$$

3.2.3.3 Index de probabilité maximale :

$$R_i = (\operatorname{argmax} P_i^j) - \frac{D}{2} \quad (3.67)$$

Kurtosis donne des valeurs élevées si le contour est bien définie. En général, les marges bien défini ont une transition abrupte le long d'une ondelette. On s'attend également que les contours bien définies ont une plus faible entropie. Index de la probabilité maximale est une autre mesure qui est déplacé pour être zéro sur la marge, positif pour l'extérieur, et négatif pour l'intérieur de la marge. Chacun de ces vecteurs a n éléments (n est le nombre de ondelettes).

Les caractéristiques suivantes peuvent être obtenues à partir de ces vecteurs.

- Moyenne et déviation standard de K
- Moyenne et déviation standard de E
- Variation totale de R :

$$\sum_i |R_i| \quad (3.68)$$

L'équation de déviation standard est :

$$\sigma = \sqrt{\frac{1}{D} \sum_{i=1}^N (P_i - \mu)^2} \quad (3.69)$$

Les caractéristiques ne sont pas dépendantes de la précision de la segmentation, parce qu'ils sont obtenus sur le long des ondelettes et si les changements des bordures par certains déflation ou inflation, ils seront les mêmes.

3.3 Les caractéristiques humaines

L'âge et la densité sont des caractéristiques humaines. La description de l'interprétation de ces traits humains est basée sur Breast Imaging Reporting and Data System (BI-RADS)[15], qui sont utilisés comme un guide pour la normalisation des rapports mammographiques, le tableau 3.2 illustre l'interprétation BI-RADS pour les valeurs de densité du sein.

Valeur de densité	Interprétation BI-RADS
1	Le sein est presque entièrement de matières grasses.
2	Il y a des densités de fibroglandulaires dispersés qui pourraient masquer une lésion sur la mammographie
3	Le sein est hétérogène dense cela peut diminuer la sensibilité de la mammographie.
4	Le tissu du sein est très dense, ce qui réduit la sensibilité de la mammographie.

TABLE 3.2 – *Interprétation BI-RADS pour les valeurs de densité du sein*

3.4 Conclusion

L'étape de l'état de l'art concernant les différents descripteurs appliqués en mammographie est nécessaire. Dans ce contexte, différents descripteurs de texture sont étudiés afin d'analyser les méthodes les plus pertinentes. Dans le cas de traitement de lésions de type masses, il est recommandé de s'appuyer sur des descripteurs de forme. Dans ce cadre, différents descripteurs de forme appliqués en mammographie ont été étudiés afin d'analyser les avantages et les inconvénients de chaque proposition. En outre, l'étude des descripteurs de contour a montré que ces derniers ont une très grande importance dans l'interprétation des masses mammaires. Cette étude détaillée nous permet de tirer profit des points forts des descripteurs existants et de tenir compte des lacunes rencontrées.

Chapitre 4

Conception et réalisation

4.1 Introduction

Malgré les efforts entrepris par les chercheurs, la classification du cancer du sein reste une tâche difficile à résoudre, la difficulté revient de plusieurs paramètres (la variabilité des images mammographiques, la variété des types de cancers qui existe, la variété de leur format).

Dans ce chapitre, nous allons exposer l'architecture de notre système dédié à la classification des masses mammaires, ce système s'inspire globalement de l'approche du médecin lors de l'examen radiologique comme c'était convenu dans le système d'aide à la rédaction des comptes rendus BI-RADS qui permet de décrire les anomalies rencontrées en mammographie

4.2 Environnement et outils de développement

4.2.1 Plateformes utilisées

Notre application a été implémentée sous Windows 8.1. Les clichés mammographiques sont décompressés sous linux elementary OS

4.2.2 Langage et environnement

Langage de programmation JAVA avec environnement " Eclipse " qui est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT (Standard Widget Toolkit), d'IBM (International Business Machines)).

Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Beaucoup d'applications et de sites Web ne fonc-

tionnent pas si Java n' est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable.

La version d'Eclipse utilisée est "Mars.1" sortie de 24 juin 2015 et est composée de 79 projets cette dernière apporte quelques nouveautés majeure :

- L'apparition d'un installateur baptisé Oomph, il permet de simplifier l'installation et le déploiement d'une version d'Eclipse avec les configurations associées au sein d'une équipe de développement.
- L'explorateur de projets peut désormais afficher des projets inclus dans d'autres projets de manière hiérarchique.
- Une partie des outils de développement a été intégrée directement sur les "update-sites".

4.2.3 La base d'images utilisées

Pour mesurer les performances de notre système de classification on a besoin d'un référentiel dans le domaine d'aide au diagnostic du cancer du sein, il existe plusieurs bases d'images : MIAS (Mammographic Image Analysis Society), DDSM (Digital Database for Screening Mammography), AMDI (Indexed Atlas of Digital Mammograms) et WDBC (Wisconsin Diagnosis Breast Cancer).

Nous avons utilisé un sous-ensemble de la base DDSM. Les types de tumeurs pris en compte dans notre système sont des masses bénigne et maligne.

DDSM : La base de données Marathon de l'université de la Floride du Sud. Une description de cette base a été effectuée par " American college of Radiology " dans le lexique de BI-RADS (Breast Imaging Reporting and Data System).

La base contient 2604 dossiers de patients. Chaque dossier de patiente est composé de :

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

L'avantage majeur de la base DDSM est qu'elle emploie le même lexique standardisé, Par

```
ics_version 1.0
filename A-1848-1
DATE_OF_STUDY 22 2 1999
PATIENT_AGE 68
FILM
FILM_TYPE REGULAR
DENSITY 2
DATE_DIGITIZED 20 5 1999
DIGITIZER HOWTEK 43.5
SEQUENCE
LEFT_CC LINES 6766 PIXELS_PER_LINE 4531 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
LEFT_MLO LINES 6436 PIXELS_PER_LINE 4396 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
RIGHT_CC LINES 6271 PIXELS_PER_LINE 4486 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
RIGHT_MLO LINES 5956 PIXELS_PER_LINE 4336 BITS_PER_PIXEL 12 RESOLUTION 43.5 OVERLAY
```

FIGURE 4.1 – *Les informations importantes dans un fichier ics*

l’American College of Radiology dans le BIRADS.

Dans cette phase des informations liées aux clichés mammographiques et aux anomalies sont extraites automatiquement à partir des fichiers .ics et .overlay, ces derniers nous offrent des informations sur les images et sur les anomalies.

4.3.1.1 Description du fichier .ics

Le fichier .ics contient des informations importantes telles que la date de l’étude, l’âge du patient, la date de la numérisation des films, le type de numériseur utilisé, une densité du tissu mammaire, la taille de chaque fichier image, le nombre de bits par pixel, la résolution de la numérisation, etc .

La figure 4.1 illustre un exemple de ce fichier.

4.3.1.2 Description du fichier overlay

Les cas anormaux ont entre un et quatre fichiers overlay, ceux-ci dépendent du nombre d’images que le radiologue marque comme sans anomalies. Pour chaque anomalie, on a des informations sur le nombre d’anomalies, le type, la forme et les bords de la tumeur présente, le degré de suspicion, le degré de subtilité, le type de pathologie et enfin la description se termine par le code du contour de l’anomalie (Figure 4.2).

Pour toutes les mammographies contenant une tumeur, le spécialiste a tracé un contour autour de la région tumorale, ce contour est analysé sous forme de chaîne à l’aide du code de Freeman. Les deux premiers chiffres de la chaîne représentent les coordonnées d’un pixel du contour respectivement la colonne et la ligne, sur lesquels on se base pour déterminer les coordonnées des autres pixels du contour.

```

TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE MASS SHAPE IRREGULAR MARGINS SPICULATED
ASSESSMENT 5
SUBTLETY 5
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 1
BOUNDARY
2248 1240 6 6 6 6 6 6 6 6 5 5 5 5 5 5 5 6 6 6 6 6 6 6
4 4 4 4 4 4 4 6 6 6 6 6 6 6 4 4 4 4 4 4 4 5 5 5 5 5

```

FIGURE 4.2 – *Un exemple d'un fichier overlay*

4.3.1.3 Conversion du LJPEG au LJPEG1

LJPEG nommé aussi JPEG-LS (souvent surnommé Lossless JPEG) est une norme de compression sans perte (donc réversible), basée sur l'algorithme LOCO-I (LOW COMplexity LOSSless COMpression for Images) et évaluée par le Joint Photographic Experts Group, dont la notoriété est reconnue pour les formats de compression JPEG ISO/CEI 10918-1 et JPEG 2000.

Dans JPEG-LS la compression est réalisée par la combinaison d'un codage adaptatif (extension des codes de Golomb) avec un codeur entropique proche du codeur de Huffman pour les zones à faible entropie. L'image est stockée étendue en tant que données binaires brutes dans un fichier qui ne contient pas d'informations d'entête. Les dimensions de l'image (hauteur et largeur) sont précisées dans le fichier "ics" pour ce cas et cette décompression est faite sous Linux.

4.3 Approche proposée

Comme nous l'avons dit plus tôt, notre but est de développer un système d'aide au diagnostic basé sur les caractéristiques utilisée dans BI-RADS[29], qui est le processus de classification utilisé par les radiologues.

Notre système se compose principalement de quatre étapes(figure4.3), l'objectif de la première étape est de faire une segmentation manuelle car on suppose que la détection est déjà faite. Le but de la deuxième étape est de quantifier les caractéristiques utilisées par les radiologistes selon BI-RADS. Dans la troisième étape parmi les caractéristiques calculées, on a sélectionne les meilleurs caractéristiques selon notre expérimentation. Et la quatrième étape nous classons les anomalies sur les deux classes (bénigne et maligne).

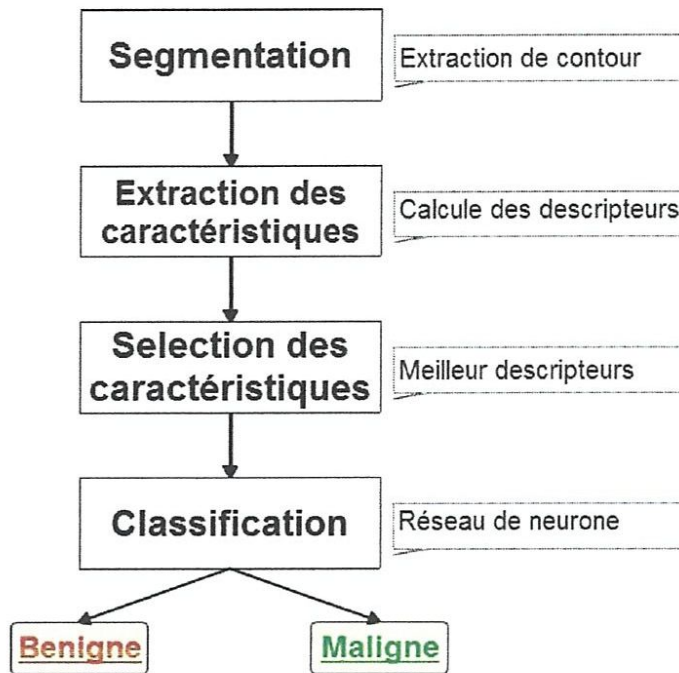


FIGURE 4.3 – Le schéma de notre application.

4.3.1 La segmentation

L'étape de la segmentation se fait manuellement par un radiologue expérimenté, parce que nous supposons que la détection de la masse est déjà faite, par conséquent, l'entrée de notre système est la région d'intérêt ROI caractérisé par une image binaire qui est composé d'une masse blanche et un fond noir.

4.3.2 Extraction des caractéristiques

Le but de cette étape est d'extraire les caractéristiques des masses mammographiques, notre système s'inspire globalement de l'approche du médecin lors d'un examen radiologique. Selon BI-RADS [4], une masse est caractérisée par des descripteurs de leur forme, de leur bord et leur densité, en plus de cela, l'âge du patient et de la taille de la masse peut reclasser une anomalie dans une classe BI-RADS inférieur ou supérieur. La figure(4.4) résume les catégories de caractéristiques utilisées par les radiologues pour classer les masses mammographiques.

4.3.2.1 Extraction des caractéristiques de forme :

Avant de commencer l'extraction des caractéristiques de forme, une étape de mise à l'échelle est nécessaire, cette phase joue un rôle très important dans le processus de calcul des caractéristiques de forme, en respectant un échelle de référence permet aux masses identiques d'avoir les mêmes "caractéristiques de forme" quelles que soient leurs tailles. Lorsqu'une masse segmentée est mise à l'échelle, l'analyse de la forme peut être réalisée.

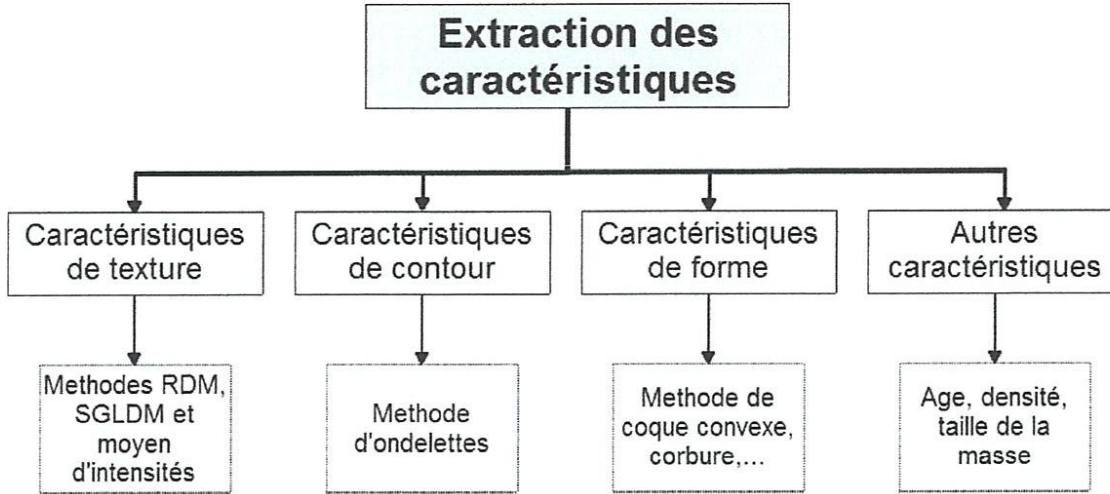


FIGURE 4.4 – Les caractéristique utilisées dans notre approche

Toutefois, plusieurs caractéristiques de forme ont été proposées dans la littérature. Dans notre travail, nous avons utilisé 10 caractéristiques :

- Courbure et continuité : Tout d’abord, un ensemble de points v_i doivent être extraites à partir du bord de la masse mammographique pour chaque changement de direction de la trajectoire de la marge, la figure 4.5 montre un exemple de processus de sélection. Les contours de la masse peuvent être représenté par un vecteur



FIGURE 4.5 – Exemple sur la sélection des points v_i .

consiste des points extraits $[v_1, v_2...v_n]$, pour chaque point i , nous discrétisons la continuité(x2) et la courbure(x1) comme suit :

$$x1 = \left| \frac{d^2 v_i}{ds^2} \right|^2 = v_{i+1} - 2v_i + v_{i+1}^2 = (x_{i-1} - 2x_i + x_{i+1})^2 + (y_{i-1} - 2y_i + y_{i+1})^2 \quad (4.1)$$

$$x2 = (x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 \quad (4.2)$$

- Compacité(x3) : La compacité est indépendant des transformations linéaires , elle mesures le degré de rugosité d’une région[54].

$$x3 = \left(\frac{2 * \sqrt{Aire * \pi}}{Perimetre} \right) \quad (4.3)$$

- Equivdiameter(EQD)(x4) : Pour différencier rondes/ovale masse de irrégulier/lobulaire

masse [54].

$$x4 = \sqrt{4 * \frac{Aire}{\pi}} \quad (4.4)$$

— Méthode de coque convexe : Nous avons mentionné dans le chapitre 3 dans la méthode de coque convexe, qu'on peut extraire 6 caractéristiques qui sont :

1. **Zone de différence(x5) :** zone de coque convexe moins la zone de la masse

2. **Moyen de variations**

$$x6 = \frac{1}{L} \sum_{i=1}^L variation(i) \quad (4.5)$$

3. **Variance de variations**

$$x7 = \frac{1}{L} \sum_{i=1}^L (variation(i) - x6)^2 \quad (4.6)$$

4. **Skewness de variations**

$$x8 = \frac{\frac{1}{L} \sum_{i=1}^L (variation(i) - x6)^3}{(\frac{1}{L} \sum_{i=1}^L (variation(i) - x6)^2)^{3/2}} \quad (4.7)$$

5. **Kurtosis de variations**

$$x9 = \frac{(1/L) \sum_{i=1}^L (variation(i) - x6)^4}{((1/L) \sum_{i=1}^L (variation(i) - x6)^2)^2} - 3 \quad (4.8)$$

6. **Entropie de variations**

$$x10 = - \sum_{i=1}^L variation(i) \log_2 variation(i) \quad (4.9)$$

Où L est le nombre de variations, un pseudo code illustrant cette technique.

Algorithme : Méthode de coque convexe

Entrée :- Image+ROI

- La coque convexe+centre de la masse

Sortie : 6 Caractéristiques de forme

Debut

1. $x5 = surface_convexe - surface_masse;$
qui représente la différence de surface de convexe et de la masse
-

2. *pour chaque 5 degré faire; la rotation du ROI chaque fois accroître avec
5 degré*
 3. $Y_min(i) = \text{clacule } Y_min(\text{ROI});$
qui représente le minimum coordonnée Y de ROI avec rotation d'angle i
 4. $Y_max(i) = \text{clacule } Y_max(\text{ROI});$
qui représente le maximum coordonnée Y de ROI avec rotation d'angle i
 5. $Variation(i) = \text{calcule variation}(Y_min, Y_max);$ *qui représente
la distance entre la projection de Y_min et Y_max sur l'axe des Y*
 6. *fin pour*
 7. $x6 = \text{calcule } \text{Moyen}[\text{Variations}];$ *qui représente le Moyen de tous
les Variations des rotations*
 8. $x7 = \text{calcule } \text{Variance}[\text{Variations}];$ *qui représente le Variance de tous
les Variations des rotations*
 9. $x8 = \text{calcule } \text{Skewness}[\text{Variations}];$ *qui représente le Skewness de tous
les Variations des rotations*
 10. $x9 = \text{calcule } \text{Kurtosis}[\text{Variations}];$ *qui représente le Kurtosis de tous
les Variations des rotations*
 11. $x10 = \text{calcule } \text{Entropie}[\text{Variations}];$ *qui représente le Entropie de tous
les Variations des rotations*
 12. $\text{Forme_Caractéristiques} = [\text{DF_AR}, \text{M_V}, \text{VR_V}, \text{SK_V}, \text{K_V}, \text{E_V}];$
 13. $\text{return_Forme_Caractéristiques};$
- Fin.*
-

4.3.2.2 Extraction des caractéristiques de contour

Comme nous l'avons déjà mentionnée dans le chapitre 3 plus précisément dans la méthode des ondelettes, on peut extraire 5 caractéristiques qui sont :

1. Kurtosis(Aplatissement) :

$$k_i = \frac{(1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^4}{((1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^2)^2} - 3 \quad (4.10)$$

sur cette caractéristique on applique la moyenne

$$x11 = \frac{1}{D} \sum_{i=1}^D P_i^j \quad (4.11)$$

et la déviation standards

$$x_{12} = \sqrt{\frac{1}{D} \sum_{i=1}^D (P_i^j - x_{11})^2} \quad (4.12)$$

2. Entropie :

$$E_i = - \sum_{j=1}^D P_i^j \log_2 P_i^j \quad (4.13)$$

La même chose pour cette caractéristique on applique la moyenne

$$x_{13} = \frac{1}{D} \sum_{i=1}^D P_i^j \quad (4.14)$$

et la déviation standards

$$x_{14} = \sqrt{\frac{1}{D} \sum_{i=1}^D (P_i^j - x_{13})^2} \quad (4.15)$$

3. Index de probabilité maximale :

$$R_i = (\operatorname{argmax} P_i^j) - \frac{D}{2} \quad (4.16)$$

on applique sur elle la variation totale

$$x_{15} = \sum_i |R_i| \quad (4.17)$$

Un pseudo-code illustre cette technique.

Algorithme : Méthode des ondelettes

Entrée :- Image+Points du contour.

- Centre de la masse.

Sortie : 5 Caractéristiques de contour

Debut

1. *pour chaque 5 points de contour faire (les ondelettes placée chaque 5 pixels)*
2. *PMF = calcule PMF(i); qui représente la probabilité de la fonction de la masse pour l'ondelette i*
3. *Ki = calcule Ki; qui représente le Kurtosis de l'ondelette i*
4. *Ei = calcule Ei; qui représente l'Entropy de l'ondelette i*
5. *Ri = calcule Ri; qui représnte l'index de maximum probabilité de l'ondelette i*
6. *fin pour*

7. $x_{11} = \text{calcule Moyenne}[k]$; qui représente le Moyenne de tous les K des ondelettes.
 8. $x_{12} = \text{calcule Standard deviation}[k]$; qui représente le Standard deviation de tous les K des ondelettes.
 9. $x_{13} = \text{calcule Moyenne}[E]$; qui représente le Moyenne de tous les E des ondelettes.
 10. $x_{14} = \text{calcule Standard deviation}[E]$; qui représente le Standard deviation de tous les E des ondelettes.
 11. $x_{15} = \text{calcule Total Variation}[R]$; qui représente la Total Variation de tous les R de ondelettes.
 12. $\text{Contour_Caractéristiques} = [M_K, S_K, M_E, S_E, TV_R]$;
 13. $\text{return Contour_Caractéristiques}$;
- Fin.*
-

4.3.2.3 Extraction de caractéristiques de texture(densité)

Dans le domaine radiologique le terme densité utilisé pour décrire le tissu de masse mammographique[29]. Les masses mammographiques ayant une haute densité sont trois fois plus susceptibles d'être maligne que les masses ayant une faible densité[57], la masse grasse peuvent être : Lipome, Fat nécrose, galactocèle ou de la graisse du sein Focal simulant une masse, en revanche la masse de densité mixte peut être ganglionnaire Fibroadenolipoma intra-mammaire[29].

Les caractéristiques de texture calculées sont :

— L'intensité de la masse(x_{16}) : Est égale au moyen d'intensités des pixels de la zone englobée par le contour.

— RDM Les caractéristiques RDM :

Les caractéristiques calculé par la méthode RDM qui sont :

1. LDE(Long Distance Emphasis)

$$x_{17} = LDE = \sum_{gdf=0}^{G-1} DGD(gdf) \cdot \ln(2/gdf) \quad (4.18)$$

2. Sharpness

$$x_{18} = Sharpness = \sum_{gdf=0}^{G-1} DGD(gdf) \cdot (gdf)^3 \quad (4.19)$$

3. SMG (Second Moment of DGD)

$$x_{19} = SMG = \sum_{gdi f=0}^{G-1} (DGD(gdi f))^2 \quad (4.20)$$

4. LDEL (Long Distance Emphasis for large Difference)

$$x_{20} = LDEL = \sum_{gdi f=0}^{G-1} DAD(gdi f)(gdi f)^2 \quad (4.21)$$

Un pseudo code illustrant cette technique :

Algorithme :RDM

Entrée :-Les directions utilisées pour la co-occurrence $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\}$
 -ROI

Sortie :Les caractéristiques calculées {LDE, Sharpness, SMG, LDEL}

Debut

1. *pour chaque angle θ faire*
2. *créer la matrice $I(\theta)$; qui représente l'image I avec une rotation (θ)*
3. *fin pour;*
4. *L = la longueur de l'image (ROI)*
5. *$MRDM = 0$; qui représente la somme de les fonctions RDM pour les 8 angles θ*
6. *pour chaque angle θ faire*
7. *$MRDM = MRDM + RDM$ de $I(\theta)$;*
8. *fin pour;*
9. *$DGD =$ calcule $DGD(MRDM)$;*
10. *$DAD =$ calcule $DAD(MRDM)$;*
11. *$x_{17} =$ calcule $LDE(DGD)$.*
12. *$x_{18} =$ calcule $Sharpness(DGD)$.*
13. *$x_{19} =$ calcule $SMG(DGD)$.*
14. *$x_{20} =$ calcule $LDEL(DAD)$.*
15. *$RDM_Caractéristiques = [LDE, Sharpness, SMG, LDEL]$;*
16. *return $RDM_Caractéristiques$;*

Fin.

— Les caractéristiques SGLDM Cette méthode utilisée pour calculer les caractéristiques désirées qui sont :

1. Corrélacion

$$X15 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad (4.22)$$

2. Homogénéité

$$X16 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad (4.23)$$

3. Variance

$$X17 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu^d)^2 \quad (4.24)$$

4. Somme moyenne

$$X18 = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad (4.25)$$

5. Prominence de cluster

$$X19 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d + j - \mu_j^d)^4 \quad (4.26)$$

6. Energie

$$X20 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad (4.27)$$

Et pour chacune de ses six caractéristiques, on applique les opérations suivantes :

1. La moyenne

$$x_{n+6} = \frac{1}{L/2} \sum_{d=1}^{L/2} x_n^d \quad (4.28)$$

2. Écart absolu moyen

$$x_{n+12} = \frac{1}{L/2} \sum_{d=1}^{L/2} |x_n^d - x_{n+6}| \quad (4.29)$$

3. Minimum

$$x_{n+18} = \text{Min}_{d=1}^{L/2} x_n^d \quad (4.30)$$

4. Maximum

$$x_{n+24} = \text{Max}_{d=1}^{L/2} x_n^d \quad (4.31)$$

5. Variance

$$x_{n+30} = \frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+6})^2 \quad (4.32)$$

6. Skewness

$$x_{n+36} = \frac{\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+6})^3}{\left(\frac{1}{L/2} \sum_{d=1}^{L/2} (x_n^d - x_{n+6})^2 \right)^{3/2}} \quad (4.33)$$

Un pseudo code illustrant cette technique

Algorithme :SGLDM

Entrée :-Les directions utilisées pour la co-occurrence $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\}$
 -ROI

Sortie :

Debut

1. *pour chaque angle θ faire*
2. *créé la matrice $I(\theta)$; qui représente l'image I avec une rotation (θ)*
3. *fin pour;*
4. *$L =$ la longueur de l'image (ROI)*
5. *pour chaque d faire ($d \in \{1, 2, 3, \dots, L/2\}$)*
6. *$Som_Co - oc = 0$; qui représente la somme de 8 co - occurrence matrices*
7. *pour chaque angle θ faire*
8. *$Som_Co - oc = Som_Co - oc + co - occurrence$ matrice de $I(\theta)$;*
9. *fin pour;*
10. *pour $i = 1$ a longueur CAC(Caractéristiques à Calculer) faire*
11. *$Si(d) =$ Calcule caractéristiques($Som_Co - oc, CAC(i)$);*
12. *fin pour;*
13. *fin pour;*
14. *pour $i = 1$ a longueur CAC faire*
15. *$Si + n = Means(Si)$;*
16. *$Si + 2n = \acute{E}cart Absolu Moyen(EAM)(Si)$;*
17. *$Si + 3n = Minimum(Si)$;*
18. *$Si + 4n = Maximum(Si)$;*
19. *$Si + 5n = Variance(Si)$;*
20. *$Si + 6n = Skewness(Si)$;*
21. *fin pour;*
22. *$SGLDM_Caractéristiques = [S]$;*
23. *return $SGLDM_Caractéristiques$;*

Fin.

4.3.2.4 Les autres caractéristiques

Nous avons utilisé dans notre système trois autres caractéristiques : la densité(x57)(le tableau 3.2 illustre l'interprétation BI-RADS pour les valeurs de densité du sein), l'âge du patient(x58) et la taille de la masse(X59). Ces deux dernières caractéristiques peuvent reclasser une anomalie dans une classe BI-RADS inférieure ou supérieure[29]. Il est facile de calculer la taille réelle de la masse en mm^2 , qui est représenté par la zone incluse dans le contour Ω . Où "res" est la résolution utilisée lors de l'étape de balayage.

$$x59 = \left(\sum_{p \subset \Omega} p \right) / (res)^2 \quad (4.34)$$

4.3.3 La sélection des caractéristiques

L'objectif principal de la sélection des caractéristiques est de réduire la dimensionnalité en éliminant les caractéristiques inutiles(insignifiantes) et la sélection des meilleures caractéristiques discriminantes.

Notre méthode de sélection est aléatoire, parmi les 59 caractéristiques calculées dans l'étape d'extraction, nous avons choisi 10 caractéristiques et on garde leurs taux de classification, après nous avons choisi 10 autres caractéristiques différentes et on garde leurs taux, ainsi de suite cette opération est répétée 100 fois. Après on a sélectionné les caractéristiques qui sont répétées dans les itérations qui ont les meilleurs taux de classification et qui sont 25 caractéristiques{Les caractéristiques de texture : moyen d'intensité, homogénéité(moyenne, écart absolu moyen), Sum-average(minimum), Energy(minimum), SMG, LDEL + tous les caractéristiques de forme et de contour + l'âge de patient, la densité et la taille de la masse}.

4.3.4 Classification

La classification dans notre approche se fait à l'aide d'un PMC(Perceptron multi Couche). PMC est un réseau neuronal artificiel avec apprentissage supervisé, à savoir, l'algorithme de rétro propagation, spécialement dédié à la classification. Il a la capacité de séparer les classes non linéairement séparables.

En fait, nous avons réalisé un système de classification(figure 4.6) consiste à classer les anomalies sur deux classes(bénignes et malignes) ce qui signifie que les neurones nombre de couche de sortie $O = 1$. Le nombre de neurones la couche d'entrée $I = 25$, il est représentée par la longueur du vecteur d'entrée. Comme mentionné dans [23] une couche cachée est suffisante pour résoudre tout problème complexe non linéairement séparable. Pour déterminer le nombre de neurones dans la couche cachée, nous nous sommes basés

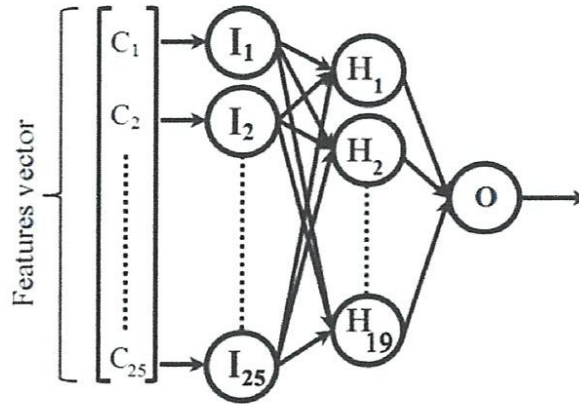


FIGURE 4.6 – L'architecture du PMC utilisée.

sur une heuristique mentionné dans[4], qui estiment que le nombre de neurones de la couche cachée H doit être inférieur à deux fois le nombre de neurones dans la couche d'entrée. Le nombre de neurones cachés H est calculé pour notre cas en utilisant la règle suivante :

$$H = O + (0.75 * I) \text{ et } H < 2 * I \quad (4.35)$$

C'est à dire : $H = 19$. Toutes les caractéristiques d'entrée ont des valeurs qui peuvent être très éloignés les uns des autres. Par exemple, la caractéristique continuité peut avoir une valeur maximale de 7000. Pour éviter cela, il faut normaliser les valeurs d'entrée en utilisant la formule suivante.

$$f_{nor} = (f - Min(f)) / (Max(f) - Min(f)) \quad (4.36)$$

Où :

- f La valeur de la caractéristique qui devrait être normalisé.
- f_{nor} Caractéristique normalisée.
- $Min(f)$ La valeur minimale de la caractéristique f .
- $Max(f)$ la valeur maximale de la caractéristique f .

Dans ce travail, nous choisissons la fonction de transfert "sigmoïde", parce que les valeurs de sortie des données utilisées sont normalisées dans l'intervalle $[0, 1]$.

4.4 Réalisation et interface

L'interface de notre application(figure 4.7) est comme suit :

1. Zone qui contient la base d'images DDSM

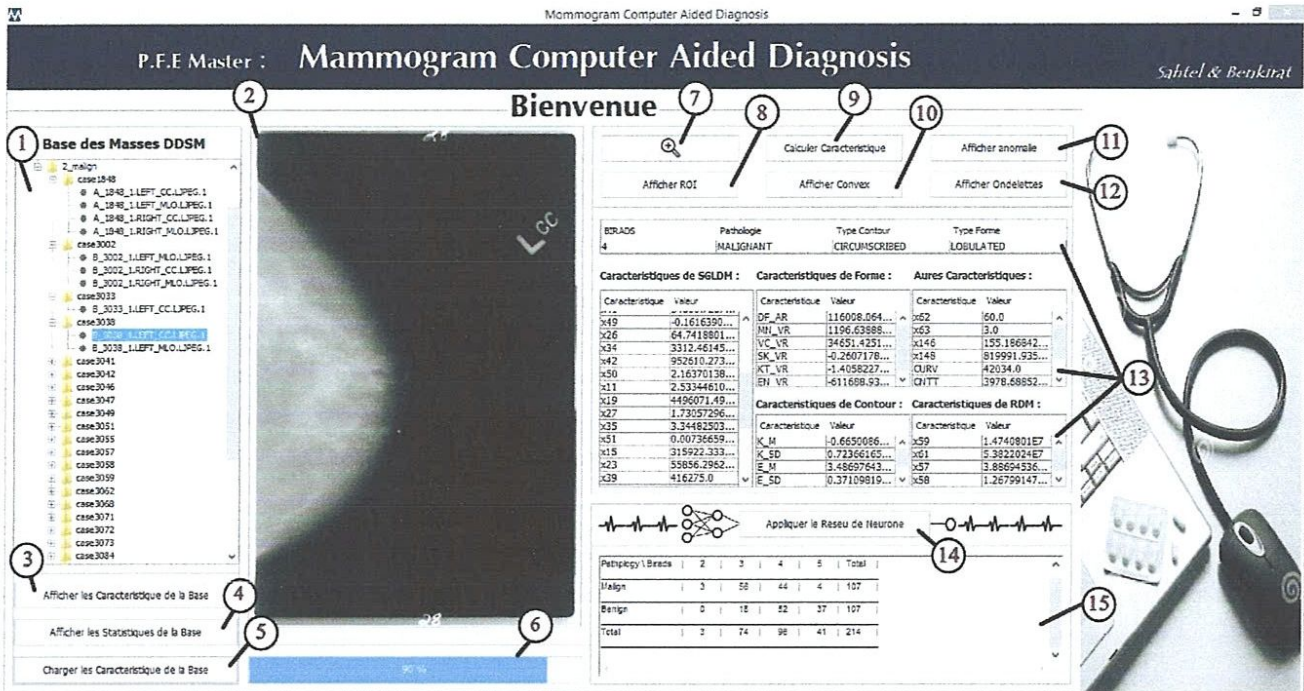


FIGURE 4.7 – Interface principale de notre application.

2. Zone pour afficher l'image sélectionnée
3. **Afficher les caractéristiques de la base** : il sert à afficher les caractéristiques de tous les clichés mammaires de la base.
4. **Afficher les statistiques de la base** : afficher les statistiques du niveau de BI-RADS ainsi que la pathologie bénigne/maligne de notre base DDSM.
5. **Charger les caractéristiques de la base** : il sert à calculer les caractéristiques des images de la base, et les mettre dans un tableau dans la base de donnée.
6. **Afficher le taux de chargement des caractéristiques**
7. **La loupe pour faire un zoom**
8. **Afficher ROI (la Région d'Intérêt)** : afficher le carré autour de la masse.
9. **Calculer caractéristique** : une fois l'image est sélectionnée, ce bouton sert à afficher toutes les caractéristiques.
10. **Afficher convexe** : afficher la forme convexe de la masse, pour clarifier l'opération

Caractéristiques	Classes BI-RADS								Pathologie		Totale
	2		3		4		5		M	B	
	M	B	M	B	M	B	M	B			
Caractéristiques de contours											
Circonscrit	0	0	0	16	2	3	0	0	2	19	21
Micro-lobulée	0	0	0	0	1	2	0	0	1	2	3
Effacée	0	0	0	3	0	6	0	0	0	9	9
Mal défini	0	0	12	4	11	18	4	0	27	22	49
Speculée	0	0	0	0	15	1	12	4	27	5	32
Caractéristiques de forme											
Ronde	0	0	0	7	1	3	0	0	1	10	11
Ovale	0	0	2	7	0	2	0	0	2	9	11
Lobulée	0	0	5	6	7	13	0	0	12	19	31
Irrégulière	0	0	5	3	21	12	16	4	42	19	61
Totale	0	0	12	23	29	30	16	4	57	57	228
	0		35		59		20		114		

TABLE 4.1 – Les statistiques de la base DDSM

de calcul des caractéristiques de forme.

11. **Afficher l'anomalie** : afficher le contour de la masse.
12. **Afficher l'ondelletes** : il sert à afficher les ondelletes autour de la masse pour clarifier l'opération de calcul des caractéristiques du contour.
13. **Zone pour afficher les caractéristiques de l'image sélectionnée**
14. **Appliquer le réseau de neurone** : pour faire une classification (bénigne ou maligne).
15. **Zone pour afficher les résultats** : dans cette zone, on a affiche quelques résultats comme le résultat du réseau de neurone, ainsi que la table des statistiques, et autres résultats.

4.5 Test et résultat

Nous avons choisi la base de données(DDSM)[40] fourni par l'université de Floride, et on a utilisé 114 masses entre maligne et bénigne, dont 57 malignes et 57 bénigne, ces masses qui sont classées en différents niveaux de BIRADS, et qui contient des différentes type

de contour(circonscrit, micro-loulée, mal défini....) et différentes forme(ronde, ovale,...) comme montre le tableau 4.1

Nous avons choisi 70 masses pour la base d'apprentissage et 44 pour le test pour évaluer notre approche de classification des masses mammographiques, il y a des paramètres qu'il faut prendre en considération comme le taux de sensibilité et le taux de spécificité.

$$sensitivité = \frac{TP}{TP + FN} \quad (4.37)$$

$$spécificité = \frac{TN}{FP + TN} \quad (4.38)$$

où TP (le nombre des masses malignes préditent comme maligne) désigne les vrais cas positifs dans les résultats de détection , et TN (le nombre des masse bénignes préditent comme bénigne) signifie les vrais cas négatifs, de même FP (le nombre des masses bénignes préditent comme maligne) représente les faux positifs cas, et FN (le nombre des masses malignes préditent comme bénigne) sont les cas de faux négatifs.

Le taux de reconnaissance est un autre paramètre très important pour l'évalutaion car il devrait être plus grand possible, ce paramètre pourrait calculé comme suit :

$$reconnaissance = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.39)$$

nous avons obtenue les résultats suivants :

- Le taux de sensibilité = 95%
- Le taux de specficite = 91.66%
- Le taux de reconnaissance= 93.18%

4.6 Conclusion

Au cours de ce chapitre, nous nous sommes étalés sur l'architecture de notre système. Nous avons commencé par donner une vue générale de ses principaux composants. L'approche sur laquelle se base notre système, afin de classifier les images mammographiques, ensuite nous avons présenté l'interface de l'application ainsi que les résultats obtenus. D'après le résultat de notre application il n y a moins de cas incorrects qui s'approuve que notre système est convaincant.

Conclusion générale

L'objectif que nous avons fixé, au début de ce mémoire, était de proposer une approche d'extraction des caractéristiques en utilisant des méthodes présentés dans le chapitre 3, et qui sont sur basées texture, forme et contour.

Pour cela nous avons utilisé la base d'images DDSM (Digital Database for Screening Mammography) pour évaluer notre approche.

Notre système se compose principalement de quatre étapes, en premier lieu nous avons fait une segmentation manuelle, ensuite on a extrait un ensemble de caractéristiques qui sont calculées à travers des méthodes connues présentés dans le chapitre 3 pour renforcer leurs pouvoir en décrire les caractéristiques pertinentes des masses mammaires. et dans la troisième étape, à travers ces caractéristiques calculées on sélectionne parmi eux les meilleurs caractéristiques. et la quatrième étape consiste à créer un réseau de neurones artificiels adéquat au problème traité afin de tester les performances de notre système.

Les résultats de la classification des données obtenues dans notre approche ont été très prometteurs.

Perspective

Comme tout travail de recherche, il existe toujours des améliorations qui peuvent lui être apportées. Pour notre travail, nous pensons d'améliorer le taux de classification (en minimisant le taux d'erreur) et la comparaison des résultats obtenus avec d'autres classificateurs.

Bibliographie

- [1] L Abid. Cancer epidemiology in algeria : best use of cancer registers. *J Afr Cancer*, 1 :98–103, 2009.
- [2] Ali Bagheri-Khaligh, Alireza Zarghami, and MT Manzuri-Shalmani. Novel margin features for mammographic mass classification. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, pages 139–144. IEEE, 2012.
- [3] Hamid Behnam, Fahimeh Sadat Zakeri, and Nasrin Ahmadinejad. Breast mass classification on sonographic images on the basis of shape analysis. *Journal of Medical Ultrasonics*, 37(4) :181–186, 2010.
- [4] Michael J Berry and Gordon Linoff. *Data mining techniques : for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.
- [5] Michel Bérubé, Belinda Curpen, Patricia Ugolini, Lucie Lalonde, and Denise Ouimet-Oliva. Level of suspicion of a mammographic lesion : use of features defined by bi-rads lexicon and correlation with large-core breast biopsy. *Canadian Association of Radiologists journal= Journal l'Association canadienne des radiologistes*, 49(4) :223–228, 1998.
- [6] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1) :1–20, 1991.
- [7] U Bottigli, D Cascio, F Fauci, B Golosio, R Magro, GL Masala, P Oliva, G Raso, and S Stumbo. Massive lesions classification using features based on morphological lesion differences. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 12, pages 20–24, 2006.
- [8] Ying Cao, Xin Hao, and Shunren Xia. An improved region-growing algorithm for mammographic mass segmentation. In *Sixth International Symposium on Multispectral Image Processing and Pattern Recognition*, pages 749710–749710. International Society for Optics and Photonics, 2009.
- [9] Heang-Ping Chan, Carl J Vyborny, HEBER MacMAHON, Charles E Metz, Kunio Doi, and Edward A Sickles. Digital mammography : Roc studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. *Investigative Radiology*, 22(7) :581–589, 1987.
- [10] Imen Cheikhrouhou, Khalifa Djemal, Dorra Sellami-Masmoudi, Nabil Derbel, and Hichem Maaref. Abnormalities description for breast cancer recognition. In *IEEE International Conference on E-medical Systems*, page na, 2007.
- [11] Chih-Yen Chen, Hong-Jen Chiou, Yi-Hong Chou, See-Ying Chiou, Hsin-Kai Wang, Szu-Yuan Chou, and Huihua Kenny Chiang. Computer-aided diagnosis of soft tis-

- sue tumors on high-resolution ultrasonography with geometrical and morphological features. *Academic radiology*, 16(5) :618–626, 2009.
- [12] David Coeurjolly, Serge Miguët, and Laure Tougne. Discrete curvature based on osculating circle estimation. In *Visual Form 2001*, pages 303–312. Springer, 2001.
- [13] Jatindra Kumar Dash and Laxmikant Sahoo. Wavelet based features of circular scan lines for mammographic mass classification. In *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*, pages 58–61. IEEE, 2012.
- [14] Joachim Dengler, Sabine Behrens, and Johann Friedrich Desaga. Segmentation of microcalcifications in mammograms. *Medical Imaging, IEEE Transactions on*, 12(4) :634–642, 1993.
- [15] Carl J D’Orsi, American College of Radiology, American College of Radiology, BI-RADS Committee, et al. *Breast Imaging Reporting and Data System : (BI-RADS)*. American College of Radiology, 1998.
- [16] B Dubuisson. Diagnostic et reconnaissance des formes, traité des nouvelles technologies, série diagnostic et maintenance. 1990.
- [17] Belal Kamal Elfarra. *Mammogram Computer-Aided Diagnosis*. PhD thesis, Islamic University Gaza Palestine, 2012.
- [18] M Le Gal, G Chavanne, and D Pellier. Valeur diagnostique des microcalcifications groupées découvertes par mammographie. *Bull cancer*, 71 :57–64, 1984.
- [19] Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5) :786–804, 1979.
- [20] Robert M Haralick, Karthikeyan Shanmugam, and Its’ Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6) :610–621, 1973.
- [21] Steven L Horowitz and Theodosios Pavlidis. Picture segmentation by a directed split-and-merge procedure. In *Proceedings of the Second International Joint Conference on Pattern Recognition*, volume 424, page 433, 1974.
- [22] IMENE CHEIKHROUHOU Epse KACHOURI. *Description et classification des masses mammaires pour le diagnostic du cancer du sein*. PhD thesis, Ph. D. Thesis. University of Evry Val d’Essone : France, 2012.
- [23] Saurabh Karsoliya. Approximating number of hidden layer neurons in multiple hidden layer bpnn architecture. *International Journal of Engineering Trends and Technology*, 3(6) :713–717, 2012.
- [24] Bertrand Kerautret and Jacques-Olivier Lachaud. Robust estimation of curvature along digital contours with global optimization. In *Discrete Geometry for Computer Imagery*, pages 334–345. Springer, 2008.
- [25] Judy Kilday, Francesco Palmieri, and Martin D Fox. Classifying mammographic lesions using computerized image analysis. *Medical Imaging, IEEE Transactions on*, 12(4) :664–669, 1993.
- [26] Jong Kook Kim, Jeong Mi Park, Koun Sik Song, and Hyun Wook Park. Adaptive mammographic image enhancement using first derivative and local statistics. *Medical Imaging, IEEE Transactions on*, 16(5) :495–502, 1997.
- [27] Marton Lanyi. Morphologic analysis of microcalcifications. In *Early Breast Cancer*, pages 113–135. Springer, 1985.
-

- [28] HEUTTE Laurent. *Reconnaissance de Formes Processus de RdF*, volume 9.
- [29] L Lévy, M Suissa, J Bokobsa, H Tristant, JF Chiche, B Martin, and G Teman. [presentation of the french translation of the breast imaging reporting system and data system (bi-rads)]. *Gynecologie, obstetrique & fertilité*, 33(5) :338–347, 2005.
- [30] Chunming Li, Chiu-Yen Kao, John C Gore, and Zhaohua Ding. Minimization of region-scalable fitting energy for image segmentation. *Image Processing, IEEE Transactions on*, 17(10) :1940–1949, 2008.
- [31] Fan Li, Mao Ye, and Xudong Chen. An extension to rough c-means clustering based on decision-theoretic rough sets model. *International Journal of Approximate Reasoning*, 55(1) :116–129, 2014.
- [32] JY Lo and CE Floyd. Predicting malignancy and invasion of breast lesions from bi-rads and history data : New results from 1027 cases. In *Radiology*, volume 217, pages 317–317. RADIOLOGICAL SOC NORTH AMER 20TH AND NORTHAMPTON STS, EASTON, PA 18042 USA, 2000.
- [33] Tomoko Matsubara, Hiroshi Fujita, Takeshi Hara, Satoshi Kasai, Osamu Otsuka, Yuji Hatanaka, and Tokiko Endo. Development of a new algorithm for detection of mammographic masses. In *Digital Mammography*, pages 139–142. Springer, 1998.
- [34] P Mohanaiah, P Sathyanarayana, and L GuruKumar. Image texture feature extraction using glm approach. *International Journal of Scientific and Research Publications*, 3(5) :1, 2013.
- [35] Naga R Mudigonda, Rangaraj M Rangayyan, and JE Leo Desautels. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *Medical Imaging, IEEE Transactions on*, 20(12) :1215–1227, 2001.
- [36] ARIF Muhammad. *Fusion de Données : Ultime Etape de Reconnaissance de Formes, Applications à l'Identification et à l'Authentification*. PhD thesis, thèse de doctorat, Université de Tours, 2005.
- [37] Thanh Phuong Nguyen and Isabelle Debled-Renneson. Curvature estimation in noisy curves. In *Computer Analysis of Images and Patterns*, pages 474–481. Springer, 2007.
- [38] Journet Nicholas. *Introduction au traitement d'images : Reconnaissance des formes*, volume 26. 2011.
- [39] American College of Radiology et al. American college of radiology breast imaging reporting and data system.
- [40] University of South Florida. Digital database for screening mammography(ddsrm), 2008.
- [41] SG Orel, DC Sullivan, and TJ Dambro. Birads categorization as a predictor of malignancy. In *Radiology*, volume 205, pages 1293–1293. RADIOLOGICAL SOC NORTH AMER 20TH AND NORTHAMPTON STS, EASTON, PA 18042, 1997.
- [42] Nikos Paragios and Rachid Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3) :223–247, 2002.
- [43] Maria Petrou and Pedro Garcia Sevilla. Image processing : dealing with texture csip. 2006.
- [44] Markus Peura and Jukka Iivarinen. Efficiency of simple shape descriptors. *Aspects of visual form*, pages 443–451, 1997.

- [45] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3) :355–368, 1987.
 - [46] Rangaraj M Rangayyan, Nema M El-Faramawy, JE Leo Desautels, and Onsy A Alim. Measures of acutance and shape for classification of breast tumors. *Medical Imaging, IEEE Transactions on*, 16(6) :799–810, 1997.
 - [47] YA Reyad, A El-Zaart, H Mathkour, M Al-Zuair, and H Al-Salman. Segmentation of fibro-glandular discs in digital mammograms using log-normal distribution. In *Recent Progress in Data Engineering and Internet Technology*, pages 475–480. Springer, 2013.
 - [48] lepage Richard. *Reconnaissance d'algues toxiques par vision artificielle et réseau de neurones*, volume 202. 2004.
 - [49] Berkman Sahiner, Heang-Ping Chan, Nicholas Petrick, Mark A Helvie, and Lubomir M Hadjiiski. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics*, 28(7) :1455–1465, 2001.
 - [50] Dorra Sellami-Masmoudi, Hichem Maaref, Imen Cheikhrouhou, Khalifa Djemal, and Nabil Derbel. Empirical descriptors evaluation for mass malignity recognition. In *First International Workshop on Medical Image Analysis and Description for Diagnosis Systems (MIAD 2009)*, pages 91–100, 2009.
 - [51] Xiangjun Shi. Mass detection and classification in breast ultrasound images. 2006.
 - [52] Radhika Sivaramakrishna, Nancy A Obuchowski, William A Chilcote, Gilda Cardenosa, and Kimerly A Powell. Comparing the performance of mammographic enhancement algorithms : a preference study. *American Journal of Roentgenology*, 175(1) :45–51, 2000.
 - [53] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.
 - [54] B Surendiran and A Vadivel. A new feature reduction method for mammogram mass classification. In *Control, Computation and Information Systems*, pages 303–311. Springer, 2011.
 - [55] Michael Unser. Description statistique de textures. 1984.
 - [56] John N Wolfe. Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, 126(6) :1130–1137, 1976.
 - [57] Ryan W Woods, Gale S Sisney, Lonie R Salkowski, Kazuhiko Shinki, Yunzhi Lin, and Elizabeth S Burnside. The mammographic density of a mass is a significant predictor of breast cancer. *Radiology*, 258(2) :417–425, 2011.
 - [58] Mingqiang Yang, Kidiyo Kpalma, and Joseph Ronsin. A survey of shape feature extraction techniques. *Pattern recognition*, pages 43–90, 2008.
 - [59] Kessentini Yousri. *Reconnaissance De Formes Classification "statistique" des formes*, volume 90. 2014.
 - [60] Dengsheng Zhang and Guojun Lu. Review of shape representation and description techniques. *Pattern recognition*, 37(1) :1–19, 2004.
-