

الجمهورية الجزائرية الديمقراطية الشعبية

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

15/888

Filière : Informatique

Option : Ingénierie Des Média

Thème :

Les caractéristiques basés contours pour la classification des masses mammographiques

Encadré Par :

Mr Ferkous Chokri

Présenté par :

Oucif Soulef

Chiheb Abir

Juin 2015

Remerciements

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 'a 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 'a 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 a tous et a toutes

Dédicace

Je dédie ce travail ...

A Mon Père ...

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation. Qu'ALLAH le tout puissant te garde toujours a mes cotés.

A Ma Mère ...

Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur.

A Ma chère sœur ...

Les mots ne suffisent guère pour exprimer l'attachement, l'amour et l'affection que je porte pour vous. Mon ange gardien et mon fidèle compagnon dans les moments les plus délicats de cette vie mystérieuse. Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite.

A Mon très cher frère ...

Mon cher petit frère présent dans tous mes moments d'examens par ton soutien moral et ses belles surprises sucrées. Je te souhaite un avenir plein de joie, de bonheur, de réussite et de sérénité. Je t'exprime à travers ce travail mes sentiments de fraternité et d'amour.

A mes chères ami (e)s : Balahcene Djamila, Atamniaa Hadjer, Oucif Soulef

Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées, vous êtes pour moi des sœurs et des amis sur qui je peux compter. En témoignage de l'amitié qui nous uni et des souvenirs de tous les moments que nous avons passé ensemble, je vous Dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

A tous les membres de ma famille, petits et grands

Abir

Dédicace

Grace à Dieu voilà notre travail terminé et il est temps pour moi de partager ma joie avec tous ceux qui m'ont soutenu.

J'ai longuement cherché l'inspiration grâce à laquelle j'allais exprimer toute ma gratitude envers ceux qui m'ont permis d'en arriver là aujourd'hui.

À travers ce modeste mémoire, je tiens à présenter mes délicates à mon père qui a consacré sa vie à notre éducation et à faire notre bonheur et qui nous encourage toujours d'achever nos études tout en espérant de voir les fruits de ses sacrifices.

À ma mère qui m'a confié de suivre mes études et m'a donné l'espoir de réussir.

Ainsi à mes frères et mes sœurs et à tous qui ont contribué de près ou de loin à la réalisation de ce travail.

À mes meilleurs amis ATAMNIA HADJER et CHIHEB ABIR qui ont toujours été là pour moi et par leur présence dans les bons comme dans les mauvais moments.

A l'ensemble des amis que j'ai connu pendant mes études et à ceux qui m'ont prodigué leurs vifs conseils, encouragements et témoigné de leur amitié.

A toute les gens qui m'aiment et qui m'ont encouragé et à ceux qui ont veillé à ce que ce travail soit à la hauteur.

C'est notamment trop touchant à dire mais les mots peuvent s'exprimer c'est juste pour ceux qui sont trop chers pour moi avec leur patience et leur amour, je les offre tout le bonheur dans cette vie et je les souhaite que dieu les garde.

Soulef

Résumé

Le cancer du sein est la maladie la plus redoutée par les femmes, par sa fréquence mais surtout son extrême gravité, Pour cette raison plusieurs études ont été proposées afin de développer des outils d'aide au diagnostic. Ces outils connus sous le nom diagnostic assistée par ordinateur qui reposent principalement sur le développement des deux phases :extraction des caractéristiques et la classification.

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 clichés mammographique très subtils provenant de la base DDSM « Digital Database for Screening Mammography »,et les résultats obtenus sont très encourageants.

Mots Clés :

Mammographie, Diagnostic assistée par ordinateur, Extraction des caractéristiques, SGLDM, RDM, La classification, Réseaux de neurones.

Abstract

Breast cancer is the most feared disease in women by its frequency but especially its extreme seriousness, for this reason several studies have been proposed to develop diagnostic tools. These tools known as computer-aided diagnosis (CADx) based primarily on development of two phases : feature extraction and classification.

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 very subtle mammograms from the Digital Database for Screening Mammography " DDSM", and the results are very encouraging.

Keywords :

Mammography, Computer-Assisted Diagnosis (CADx), Features extraction, SGLDM, RDM, Classification, Neural network.

Table des matières

Introduction générale	14
1 Intelligence artificiel et reconnaissance de formes	17
1.1 Introduction	18
1.2 Définition de RDF	18
1.3 Principe des RDF	19
1.4 Processus de RDF	19
1.4.1 monde physique	20
1.4.2 codage	20
1.4.3 Pré traitements	20
1.4.4 Extraction des caractéristiques	20
1.4.5 Apprentissage	21
1.4.5.1 Apprentissage supervisé	21
1.4.5.2 Apprentissage non supervisé	21
1.4.6 Décision	21
1.5 Classification en RDF	22
1.5.1 Approche conventionnelle	22
1.5.1.1 Approche statistique	22
1.5.1.1.1 Méthodes paramétriques	22
1.5.1.1.2 Méthodes non paramétriques	22
1.5.1.2 Approche structurale	23
1.5.1.2.1 Arbre de décision	23
1.5.1.2.2 Algorithme de base	23
1.5.2 Approche fondée sur l'intelligence Artificielle	24
1.5.2.1 La classification par réseau de neurone	24
1.5.2.1.1 La connectivité	25
1.5.2.1.2 Le perceptron	25
1.5.2.1.3 Le perceptron multicouches	27
1.6 Mesure de performance	28
1.7 Domaine d'application	28
1.8 Conclusion	30

2	Extraction des caractéristiques des masses mammaires	31
2.1	Introduction	32
2.2	Description	33
2.2.1	Les descripteurs de texture en mammographie	34
2.2.1.1	Les statistiques de premier ordre	35
2.2.1.2	La matrice de cooccurrence	36
2.2.1.3	Difference Matrix (RDM)	39
2.2.1.4	Gray Level Co-occurrence Matrix (GLCM)	41
2.2.1.5	La transformée de Fourier	42
2.2.1.6	L'analyse fractale	43
2.2.2	Les descripteurs de forme en mammographie	44
2.2.2.1	Les descripteurs géométriques	45
2.2.2.1.1	Le centre de gravité	45
2.2.2.1.2	L'aire	46
2.2.2.1.3	Le périmètre	40
2.2.2.1.4	La circularité	46
2.2.2.1.5	La rectangularité	47
2.2.2.1.6	La compacité	48
2.2.2.2	Les descripteurs spécifiques	48
2.2.2.2.1	La courbure	48
2.2.2.2.2	Le nombre des protubérances et des dépressions importantes	49
2.2.2.2.3	Le squelette elliptique normalisé	50
2.2.2.2.4	La longueur radiale normalisée	51
2.3	Conclusion	54
3	Cancer du sein et pathologies mammaires	55
3.1	Introduction	56
3.2	Le cancer du sein	56
3.3	La Mammographie	57
3.4	Les pathologies mammaires	58
3.4.1	Les micros calcifications	58
3.4.1.1	Morphologie	59
3.4.1.2	Distribution	61
3.4.2	Les masses	63
3.4.2.1	La forme	63
3.4.2.2	Le contour	64
3.4.2.3	La densité	65
3.4.3	La distorsion architecturale	66
3.4.4	Les cas particuliers	66
3.4.5	Les signes associés	66

3.5	La classification des pathologies mammaires	66
3.5.1	La classification de Legal	66
3.5.2	La classification de Lanyi	68
3.5.3	La classification de Frouge	68
3.5.4	La classification de Moskowitz	69
3.5.5	La classification de Sigfusson	69
3.5.6	La classification BIRADS	70
3.6	Conclusion	73
4	Conception et Application	74
4.1	Introduction	75
4.2	Environnement et outils de développement	75
4.2.1	Plateformes utilisées	75
4.2.2	Langage et environnement	75
4.2.3	Choix du SGBD	77
4.2.4	Java PersistenceAPI (JPA)	77
4.3	Description générale du système	78
4.3.1	Gestion de la base DDSM	80
4.3.1.1	Description de fichier ics	81
4.3.1.2	Description du fichier overlay	81
4.3.1.3	Le modèle conceptuel des données (MCD)	82
4.3.1.4	Conversion du LJPEG au LJPEG1	83
4.3.2	Calcul des caractéristiques	84
4.3.2.1	Les caractéristique RDM (RunDifference Matrix)	84
4.3.2.2	Les caractéristiquesSGLDM (Spatial Gray Level Dependence Method)	85
4.3.2.3	Les Caractéristiques humain	86
4.3.2.4	Autres caractéristiques	87
4.3.3	Développement d'un RNA	87
4.3.3.1	Architecture générale du RNA	88
4.3.3.2	La couche de sortie	88
4.3.3.3	La couche d'entrée	88
4.3.3.4	La couche cachée	88
4.4	Réalisation et interface	89
4.5	Résultats et discussion	93
4.6	Conclusion	95
	Conclusion générale	96
	Bibliographie	97

Table des figures

1.1	Processus de RDF.	19
1.2	Arbre de décision.	24
1.3	perceptron mono couche.	25
1.4	les formes de fonction de transfert	26
1.5	perceptron multi couche à deux couche cachée	27
2.1	Principales méthodes de caractérisation de masses mammaires.	33
2.2	La cooccurrence pour un pixel (jaune) avec $d = 3$ pixels et $\theta \in \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}\}$	36
2.3	Run Difference Matrix (RDM) is a function of r and level difference with the given direction θ	39
2.4	Relation entre complexité du contour et malignité :a)masse circonscrite circulaire, b) circonscrite ovale, c) macro lobulée, d) micro lobulée, e)irrégulière, f) spéculé.	44
2.5	Calcul de la mesure de circularité.	47
2.6	Calcul de la mesure de rectangularité.	47
3.1	Les composants d'un mammographe.	58
3.2	Les différents types de microcalcifications : a) cutanées ou dermiques, b) vasculaires,c) grossières ou coralliformes, d) en bâtonnets, e) rondes, f) a centres clairs, g)en coquille d'oeuf ou pariétales, h) a type de lait clacique, i) de suture, j) dystrophiques, k) amorphes ou indistinctes, l) fines polymorphes, m) linéaires ramifiées	60
3.3	Les différentes distributions des microcalcifications : a) Mcs diffuses ou éparses,Répartition : b) Régionale, c) En cluster, d) Linéaire et e) Segmentaire.	62
3.4	Les différentes formes possibles d'une masse : a) Ronde, b) Ovale, c) Lobulée , d) Irrégulière	63
3.5	Les différents contours possibles d'une masse : a) Circonscrit, b) Microlobulé,c) Masqué, d) Indistinct et e) Spiculé	64

3.6	- Densité mammaire selon le lexique BIRADS : a) Densité < 25 %, b) Densité entre 25 et 50%, c) Densité entre 50 et 75% et d) Densité > 75%.	65
4.1	Démarrage d'Eclipse.	76
4.2	Une application qui utiliserait une classe Client persistante. . .	77
4.3	Schéma générale de notre approche.	79
4.4	Les informations importantes dans un fichier ics.	81
4.5	Un exemple d'un fichier overlay.	82
4.6	Le modèle conceptuel des données.	83
4.7	Le code de décompression d'images Ljpeg.	84
4.8	Un pseudo code de la méthode RDM.	85
4.9	Un pseudo code de la méthode SGLDM.	86
4.10	interface principale de notre application.	89
4.11	Paramètres d'ajustement.	90
4.12	Chargement des caractéristiques d'apprentissage.	91
4.13	Chargement des caractéristiques de teste.	91
4.14	Affichage d'anomalie	92
4.15	Affichage de ROI	93
4.16	Affichage de résultat	94

Liste des tableaux

2.1	Les avantages et les inconvénients des descripteurs géométriques et spécifiques	53
3.1	La classification de Moskowitz.	69
3.2	La classification de Sigfusson.	70
3.3	Classification BIRADS de l'ACR.	72
3.4	Conduite a tenir pour chaque classe de l'ACR.	73
4.1	Interprétation BIRADS pour les valeurs de densité du sein . . .	86

Introduction générale

Au cours des dernières années, Le cancer du sein représente l'un des enjeux prépondérants dans le domaine de la santé publique qui a commencé à réclamer de l'importance dans la recherche médicale en raison de sa prévalence élevée et les taux de détection depuis les dernières décennies.

Selon les statistiques de l'institut national de la santé publique Algérienne (INSP) [43], le nombre de décès annuels dus à cette maladie est évalué à 3500. En outre, on recense 7000 nouveaux cas de maladie par an. L'incidence a été multipliée par 5 en 20 ans. La prévalence a grimpé de 10 pour 100000 habitants en 1990 à 50 pour 100000 habitants actuellement.

La prévention du cancer du sein jusqu'à présent n'est pas une solution fiable, cela nécessite qu'une détection précoce soit obligatoire afin de réduire ou d'empêcher la perte de vie due au cancer du sein.

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. Elle permet de mettre en évidence des anomalies telles que les opacités et les calcifications qui peuvent traduire des lésions malignes. 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 pré traitement d'image, que l'idée d'utiliser la machine pour faciliter le travail du radiologue pour mieux diagnostiquer un cliché mammographique. La recherche menée dans ce domaine actuellement est encore à un stade peu avancé. Pour l'instant, seul le problème de la détection des micro-calcifications a prouvé sa faisabilité. Les résultats plus qu'encourageants obtenus dans ce domaine ont permis d'espérer un aboutissement aux autres problèmes plus complexes comme celui des masses.

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'extraction 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 :intelligence artificielle et la reconnaissance de forme.

La reconnaissance des formes vise la perception et l'interprétation de l'environnement et qui est une phase préliminaire qui consiste à numériser des éléments du monde réel (acquisition des données).

Ce premier chapitre consiste à donner une vue générale sur l'intelligence artificiel et la reconnaissance des formes, on décrit le processus de reconnaissance et les méthodes statistiques et structurelle et à la fin quelques domaine d'où la reconnaissance est utilisée.

Chapitre 2 :Extraction des caractéristiques des masses mammaires.

On se focalise à la partie description et l'extraction des caractéristiques des masses mammaires .On commence par aborder les descripteurs de texture les plus fréquents tels queLes statistiques de premier ordre, la matrice de cooccurrence,différence matrice (RDM), la transformée de FourierEt l'analyse fractale.

Ensuite, on se focalise sur les différents descripteurs de forme étantDonné qu'ils s'avèrent plus performants que ceux de texture dans le cas de lésions de typeMasse. L'étude de ces descripteurs permet d'identifier les avantages et les inconvénients de chaque méthode .

Chapitre 3 :Cancer du sein et pathologies mammaires.

Le cancer du sein est défini afin de justifier l'importance accordée à ce type de cancer. Une orientation vers la mammographie. Puis, nous détaillons les caractéristiques (morphologie, distribution, forme, contour, ...) des lésions mammaires qui sont les micros 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 les différentes classifications des lésions mammaires telles que la classification de Legalet celle de BIRADS. On adopte dans ce travail la classification de BIRADS qui est la plus récente et qui tient en compte les lacunes des autres classifications présentées dans la littérature.

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 la base d'images utilisé et ses composants et les caractéristiques calculées.

Chapitre 1

Intelligence artificiel et reconnaissance de formes

Sommaire

1.1	Introduction	18
1.2	Définition de RDF	18
1.3	Principe des RDF	19
1.4	Processus de RDF	19
1.4.1	monde physique	20
1.4.2	codage	20
1.4.3	Pré traitements	20
1.4.4	Extraction des caractéristiques	20
1.4.5	Apprentissage	21
1.4.6	Décision	21
1.5	Classification en RDF	22
1.5.1	Approche conventionnelle	22
1.5.2	Approche fondée sur l'intelligence Artificielle	24
1.6	Mesure de performance	28
1.7	Domaine d'application	28
1.8	Conclusion	30

Chapitre 1

INTELLIGENCE ARTIFICIEL ET RECONNAISSANCE DE FORMES

1.1 Introduction

La reconnaissance de formes est une capacité commune à tous les organismes vivants. Cependant, les différentes créatures reconnaissent de manières différentes. Un homme peut reconnaître une autre personne par son visage, par sa voix ou bien par son écriture manuscrite etc...mais un chien reconnaît une personne ou un autre chien par son odorat. Un aveugle reconnaît des objets en les touchant. Tous ces exemples sont à classer dans le domaine de la reconnaissance.

L'homme est un système biologique le plus parfait des systèmes de reconnaissance de formes . Dans le but est de concevoir des systèmes capables de percevoir et d'interpréter des informations captées dans le monde réel.

La reconnaissance de formes est un domaine de recherche très actif ayant une vaste diversité d'applications réelles.

Dans ce chapitre, nous décrivons premièrement les techniques les plus utilisées dans la reconnaissance de formes par la suite nous présentons les outils utilisés et proposés pour la classification des formes dans la littérature.

1.2 Définition de RDF

La reconnaissance de formes 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 variées.

Il peut s'agir de contenu visuel (code barre, visage, empreinte digitale...) ou sonore (reconnaissance de parole), d'images médicales multispectrales (images satellitaires) ou (rayon X, EEG, IRM...) et bien d'autres. La reconnaissance de formes s'intéresse à la conception et à la réalisation de systèmes (matériels et logiciels) capables de percevoir, et dans une certaine mesure, d'interpréter des signaux captés dans le monde physique [25].

1.3 Principe des RDF

La reconnaissance des formes (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. Il s'agit donc d'un outil qui a entre autres, la capacité d'apprentissage [26].

1.4 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) [27].

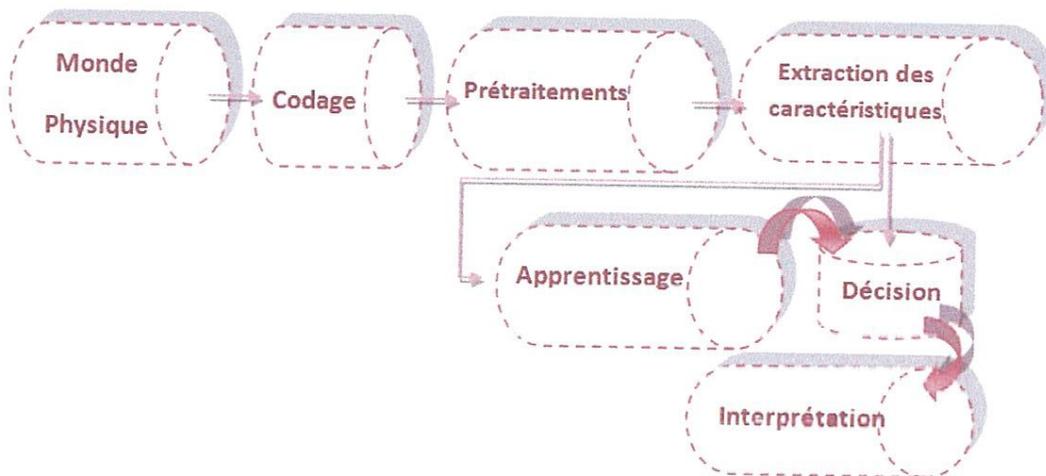


FIGURE 1.1 – Processus de RDF.

1.4.1 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 [28].

1.4.2 codage

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

1.4.3 Pré traitements

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.

1.4.4 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 cooccurrence, des descripteurs de Fourier, du spectre

de puissance, des moments, aussi bien que de diverses primitives structurales.

- 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 la modélisation AR, de fonctions d'approximation, des zero crossing, et de plusieurs types de segments de traits structurels[29].

1.4.5 Apprentissage

L'apprentissage tente de définir des classes de décision ou d'appartenance. Son rôle est déclarer la décision à l'aide de connaissance à priori sur les formes, à partir de critères spécifiques aux formes. il existe 2 type d'apprentissage (supervisé , non supervisé)[34].

1.4.5.1 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 [33] .

1.4.5.2 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 [33].

1.4.6 Décision

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

1.5 Classification en RDF

Les approches en RDF peuvent être classées en deux catégories : l'approche conventionnelle et l'approche fondée sur l'intelligence Artificielle. Les techniques conventionnelles relèvent de deux méthodologies : statistique et structurelle. Les techniques basées sur l'intelligence Artificielle permettent de représenter et de prendre en compte diverses sources de connaissances qui sont représentées par un formalisme bien spécifique.

1.5.1 Approche conventionnelle

1.5.1.1 Approche statistique

La statistique et l'analyse de données sont parmi les premières disciplines à s'être intéressées aux problèmes de classification. Ces disciplines ont étudié et proposé de multiples méthodes de classification, comme par exemple la régression, le modèle linéaire, l'analyse discriminante, ... etc.

Cependant, certaines de ces techniques imposent des contraintes difficilement interprétables par l'utilisateur qui n'est en général ni un spécialiste des mathématiques ni des statistiques.

En générale cette approche consiste à l'extraction des caractéristiques produit des vecteurs de paramètres qui sont confrontés à des modèles numériques caractérisant chaque classe [30].

1.5.1.1.1 Méthodes paramétriques

Pour chaque classe, il faut estimer les paramètres de la fonction de discrimination. Le choix de la fonction de discrimination dépend de la nature des distributions des classes et de leur bruit [31].

La distribution des classes dépend du choix de caractéristique. La classification paramétrique exige que les échantillons des classes soient regroupés dans les régions disjointes de l'espace de caractéristiques.

1.5.1.1.2 Méthodes non paramétriques

Les méthodes non paramétriques génèrent les fonctions d'appartenance des classifieurs [32].

Dans le cas de ces méthodes, les lois de probabilité sont inconnues pour une des classes. Le problème revient à établir des frontières de décision entre les classes.

1.5.1.2 Approche structurelle

Les formes sont représentées par des structures complexes composées de primitives simples et de relation topologique entre les primitives, en reconnaissance de forme structurelle. Les relations et les primitives sont caractérisées par des attributs symboliques extraits des données.

On peut dire aussi qu'une forme est représentée par une décomposition en sous-parties et par la description de la relation entre ces sous parties [42].

1.5.1.2.1 Arbre de décision

Les arbres de décision sont des règles de classification qui basent leur décision sur une suite de tests associés aux attributs, les tests étant organisés de manière arborescente, leurs principe est de classer les objets en sous classes par division hiérarchique (construction automatique à partir de l'échantillon de la base). Décomposer l'espace des caractéristiques selon la variable la plus discriminante à chaque étape.

1.5.1.2.2 Algorithme de base

1. choisir le meilleur attribut.
2. étendre l'arbre en rajoutant une nouvelle branche pour chaque valeur de l'attribut.
3. Répartir les exemples d'application sur les feuilles de l'arbre.
4. Si aucun exemple n'est mal classé alors arrêt, sinon répéter les étapes 1-4 pour les feuilles.

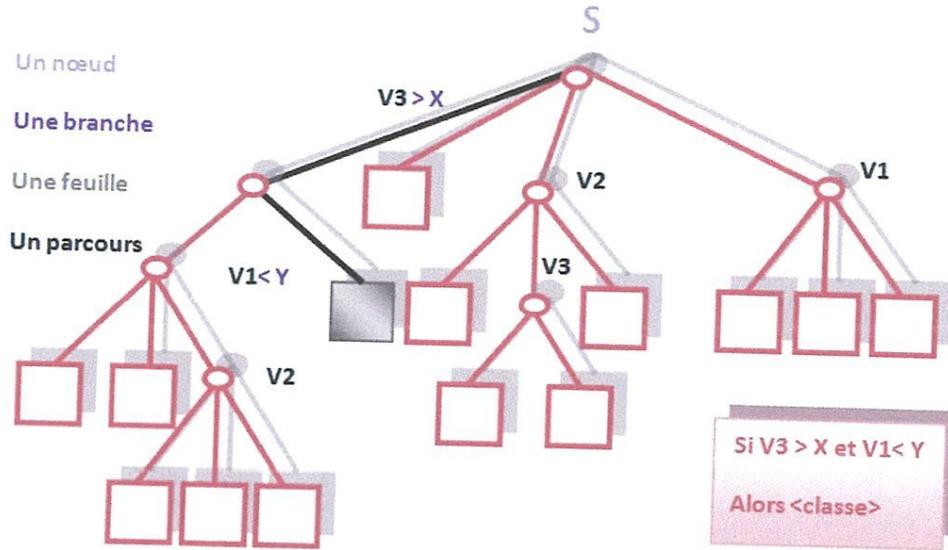


FIGURE 1.2 – Arbre de décision.

1.5.2 Approche fondée sur l'intelligence Artificielle

1.5.2.1 La classification par réseau de neurone

Un réseau de neurone est un modèle de calcul dont la conception est schématiquement inspirée du fonctionnement de vrais neurones humaines. Cette technique est placée dans la famille des méthodes de l'intelligence artificielle qu'ils enrichissent en permettant de prendre des décisions en s'appuyant d'avantage sur la perception que sur le raisonnement logique formel.

Un réseau de neurones artificiels est un ensemble de neurones formels, connectés entre eux selon une certaine architecture (topologie) [35]. Et ils sont généralement assimilés à l'intelligence artificielle.

Haykin [36] en propose la définition suivante :

Un réseau de neurones est un processus distribué de manière massivement parallèle, qui a une propension naturelle à mémoriser des connaissances de façon expérimentale et de les rendre disponibles pour utilisation. Il ressemble au cerveau en deux points :

- la connaissance est acquise au travers d'un processus d'apprentissage ;
- les poids des connections entre les neurones sont utilisés pour mémoriser la connaissance.

La plupart des réseaux de neurones ont une certaine capacité d'apprentissage. Cela signifie qu'ils apprennent à partir d'exemples, de même que les enfants apprennent à distinguer les chiens des chats à partir d'exemples de chiens et de chats.

1.5.2.1.1 La connectivité

La connectivité des réseaux est la manière dont les neurones sont connectés entre eux. Elle peut être totale (tous les neurones sont connectés entre eux) ou organisée par couche (les neurones d'une couche ne sont connectés jusqu'à la couche précédente en entrée et à la couche suivante en sortie). Chaque neurone réalise une opération très simple, qui est en fait une somme pondérée de ses entrées[37]. Le résultat est comparé à un seuil et le neurone devient excité si ce seuil est dépassé. Il existe des réseaux mono couches ou multi couches.

1.5.2.1.2 Le perceptron

C'est un des premiers réseaux de neurones, conçu en 1958 par Rosenblatt. Il est linéaire et mono couche. Le perceptron est l'un des réseaux de neurones les plus utilisés pour des problèmes d'approximation, de classification et de prédiction. La représentation graphique (conventionnelle) d'un neurone formel modélisé par Mc Culloch et Pitts (voir la figure).

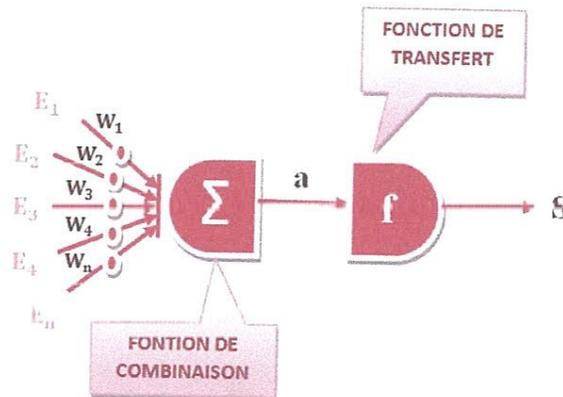


FIGURE 1.3 – perceptron mono couche.

- Les entrées "E" du neurone proviennent soit d'autres éléments "processeurs", soit de l'environnement.
- Les poids "W" déterminent l'influence de chaque entrée.
- La fonction de combinaison "a" combine les entrées et les poids.
- La fonction de transfert calcule la sortie "S" du neurone en fonction de la combinaison en entrée [38].
- La Fonction de Combinaison calcule l'influence de chaque entrée en tenant compte de son poids. Elle fait la somme des entrées pondérées :

$$a = \sum W_i E_i \quad (1.1)$$

- W_i : Poids de la connexion à l'entrée i
- E_i : Signal de l'entrée i.
- La Fonction de Transfert détermine l'état du neurone (en sortie).
- $S = f(a)$ ou encore :

$$S = f\left(\sum W_i E_i\right) \quad (1.2)$$

- La fonction de transfert "f" peut avoir plusieurs formes(voir la figure).

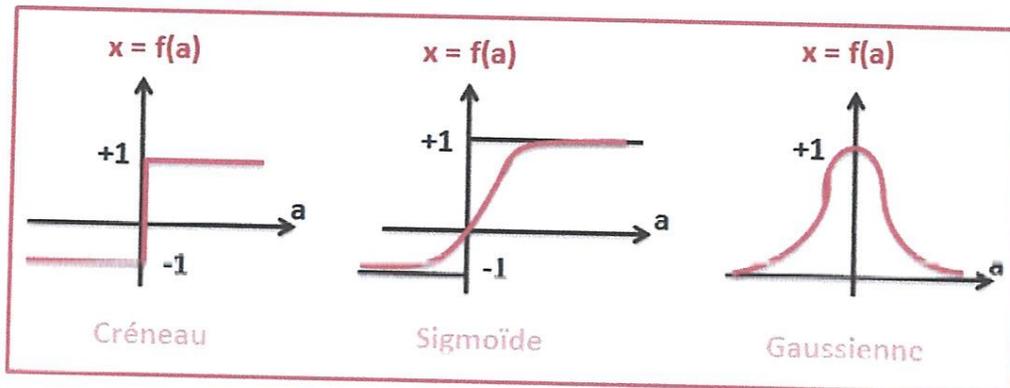


FIGURE 1.4 – les formes de fonction de transfert .

1.5.2.1.3 Le perceptron multicouches

La mise en cascade de perceptrons conduit à ce qu'on appelle le perceptron multicouches.

Le Perceptron multicouche est un classifieur linéaire de type réseau neuronal formel organisé en plusieurs couches au sein desquelles une information circule de la couche d'entrée vers la dernière couche uniquement [39].

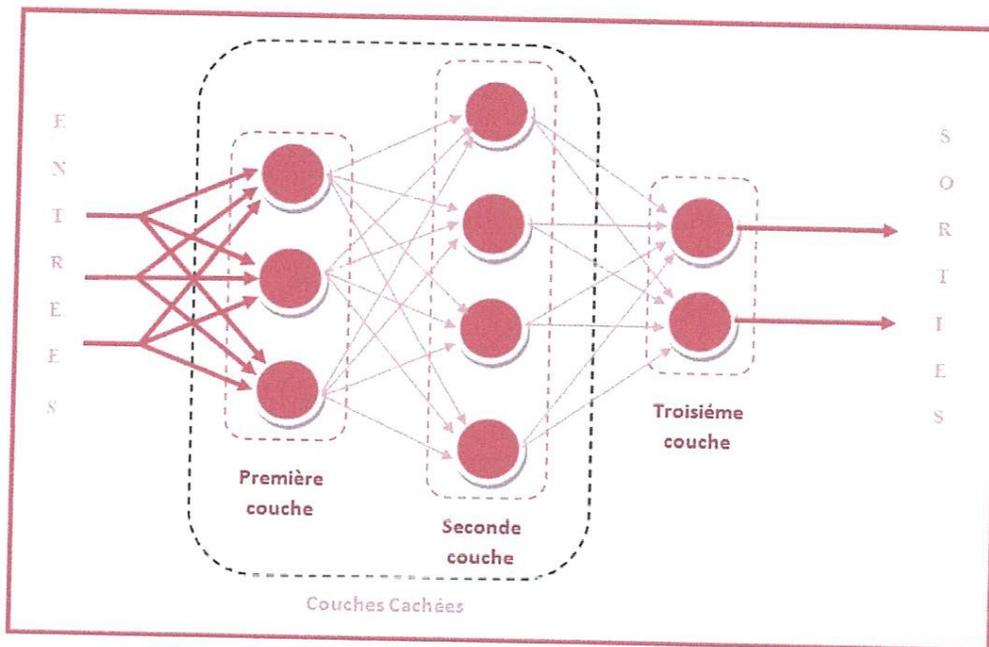


FIGURE 1.5 – perceptron multi couche à deux couche cachée .

La dernière couche du réseau est appelée couche de sortie, les autres étant désignées sous le terme de couches cachées car les valeurs de sortie de leurs neurones ne sont pas accessibles de l'extérieur.

Dans le cas général, un perceptron multicouche peut posséder un nombre de couches quelconque et un nombre de neurones (ou d'entrées) par couche également quelconque [40].

L'avantage de ce réseau par rapport au perceptron monocouche est que la limite des classes qu'il sait reconnaître devient convexe dans l'espace des motifs, s'il ne comporte qu'une couche cachée et quelconque s'il en comporte

au moins deux. Ce type de réseau est l'un des plus employés actuellement car ses performances sont généralement bonnes.

1.6 Mesure de performance

Pour mesurer l'efficacité d'un système de reconnaissance univoque on utilise :

- le taux de reconnaissance : le nombre de décisions correctes divisé par le nombre de décisions total.
- le taux de rejet : le nombre de décisions non univoques divisé par le nombre de décisions total.
- le taux d'erreur : le nombre de décisions univoques erronées divisé par le nombre de décisions total.
ou **taux de reconnaissance = 1 - taux de rejet - taux d'erreur**
[41]

1.7 Domaine d'application

- **Robotique/industrie**

- Assemblage (reconnaissance de pièces)
- Contrôle de qualité.
- Véhicule autonome, etc.

- **Téledétection**

- Météo (tempête, ouragan...)
- Identification et suivi des cultures, des forêts, des réserves d'eau.
- Cartographie.

1.5.2.1.3 Le perceptron multicouches

La mise en cascade de perceptrons conduit à ce qu'on appelle le perceptron multicouches.

Le Perceptron multicouche est un classifieur linéaire de type réseau neuronal formel organisé en plusieurs couches au sein desquelles une information circule de la couche d'entrée vers la dernière couche uniquement [39].

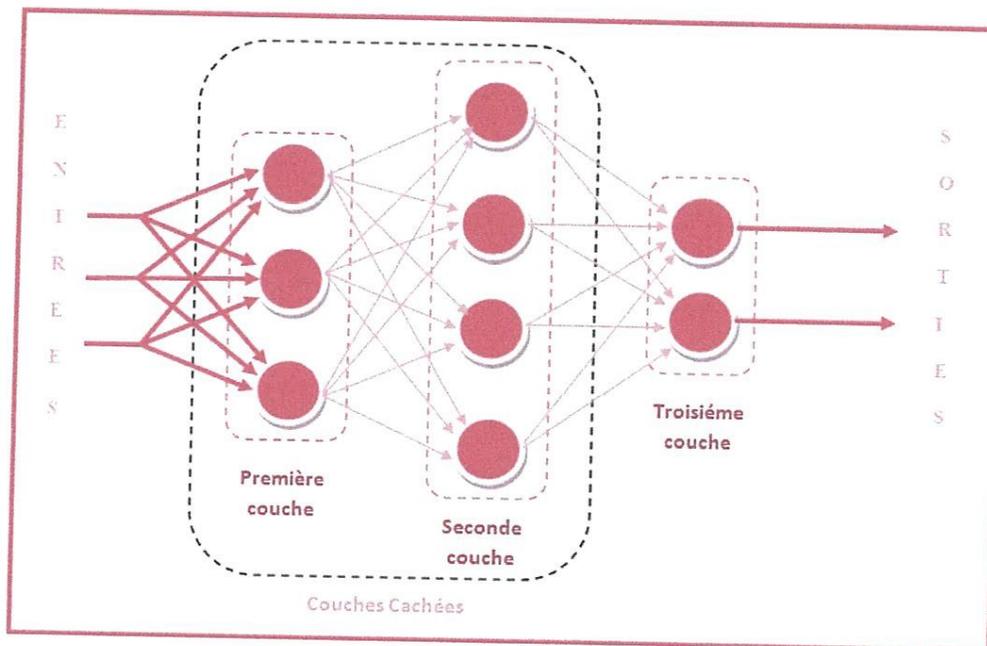


FIGURE 1.5 – perceptron multi couche à deux couche cachée .

La dernière couche du réseau est appelée couche de sortie, les autres étant désignées sous le terme de couches cachées car les valeurs de sortie de leurs neurones ne sont pas accessibles de l'extérieur.

Dans le cas général, un perceptron multicouche peut posséder un nombre de couches quelconque et un nombre de neurones (ou d'entrées) par couche également quelconque[40].

L'avantage de ce réseau par rapport au perceptron monocouche est que la limite des classes qu'il sait reconnaître devient convexe dans l'espace des motifs, s'il ne comporte qu'une couche cachée et quelconque s'il en comporte

au moins deux. Ce type de réseau est l'un des plus employés actuellement car ses performances sont généralement bonnes.

1.6 Mesure de performance

Pour mesurer l'efficacité d'un système de reconnaissance univoque on utilise :

- le taux de reconnaissance : le nombre de décisions correctes divisé par le nombre de décisions total.
- le taux de rejet : le nombre de décisions non univoques divisé par le nombre de décisions total.
- le taux d'erreur : le nombre de décisions univoques erronées divisé par le nombre de décisions total.
ou **taux de reconnaissance = 1 - taux de rejet - taux d'erreur**
[41]

1.7 Domaine d'application

- **Robotique/industrie**

- Assemblage (reconnaissance de pièces)
- Contrôle de qualité.
- Véhicule autonome, etc.

- **Téledétection**

- Météo (tempête, ouragan...)
- Identification et suivi des cultures, des forêts, des réserves d'eau.
- Cartographie.

- Analyse des ressources terrestres.
- Pollution.
- **Application militaire**
 - Guidage de missile (reconnaissance d'une cible et du terrain).
 - Reconnaissance aérienne (espionnage).
- **Bureautique**
 - Reconnaissance de texte par ordinateur (OCR).
 - Analyse de document.
 - Reconnaissance de la parole.
- **Sécurité**
 - Identification des empreintes digitales (iris, main).
 - Reconnaissance de visage.
 - Authentification de la parole.
 - Identification de signature.
- **Médecine**
 - Échographie, IRM (Imagerie par Résonance Magnétique), Microscope (histologie) etc. pour détecter des tumeurs, cellules cancéreuses ou autres maladies.
 - Analyse de l'ECG à des fins de diagnostic.
 - Analyse d'images médicales.

1.8 Conclusion

La reconnaissance de formes est un domaine de l'informatique destiné à l'étude et à la conception de module de reconnaissance dans les systèmes qui visent à automatiser certaines tâches de l'homme.

Le choix d'une méthode de classification dans le domaine de la reconnaissance de formes pour un problème donné est très difficile parce que toutes les méthodes proposées pendant des années ont des points forts et faibles.

Dans ce chapitre, nous avons présenté les notions fondamentales de la reconnaissance de formes.

Chapitre 2

Extraction des caractéristiques des masses mammaires

Sommaire

2.1	Introduction	32
2.2	Description	33
2.2.1	Les descripteurs de texture en mammographie . . .	34
2.2.2	Les descripteurs de forme en mammographie . . .	44
2.3	Conclusion	54

Chapitre 2

EXTRACTION DES CARACTÉRISTIQUES DES MASSES MAMMAIRES

2.1 Introduction

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 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.

La description des lésions mammaires est une étape incontournable dans la chaîne de traitement des images mammographies. En effet, les descripteurs sont les entités qui vont reformuler l'aspect pathologique des lésions en entités mathématiques. D'où, plus les descripteurs traduisent au mieux la description de la lésion, plus le résultat de classification est précise.

Les images mammographies sont présentées en niveau de gris et ne contiennent pas d'information couleur. De ce fait, on se contente à étudier les descripteurs de texture et de forme. 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.

2.2 Description

On peut distinguer différentes méthodes de caractérisation de la forme des lésions étudiées : descripteurs de texture, descripteurs de forme. Nous avons essayé de proposer une classification de ces méthodes selon le schéma suivant (voir la figure) :

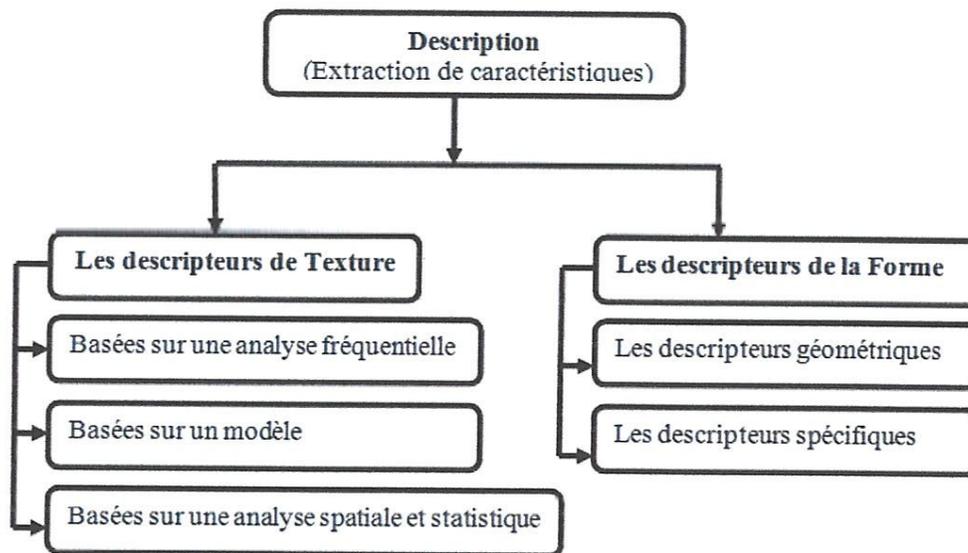


FIGURE 2.1 – Principales méthodes de caractérisation de masses mammaires.

2.2.1 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 [1] 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 un thèse, Unser [2] 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 au travers de celle-ci se traduit par une perception (impression) visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée ».

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 cooccurrence .

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

2.2.1.1 Les statistiques de 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 :

$$HIST(i) = \frac{1}{ng} \sum_{p=1}^{ng} I(p) \quad (2.1)$$

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 :

$$Moy = \frac{1}{ng} \sum_{i=1}^{ng} i.Hist(i) \quad (2.2)$$

– type L'écart :

$$\sigma - Hist = \sqrt{\frac{1}{ng} \sum_{i=1}^{ng} (i - Moy)^2.Hist(i)} \quad (2.3)$$

– La variance :

$$Var = \frac{1}{N} \sum_{p=1}^N (I(p) - Moy)^2 \quad (2.4)$$

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 s'avère plus efficace.

2.2.1.2 La matrice de cooccurrence

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 mammographies est la méthode de matrice de cooccurrence ou méthode de dépendance spatiale des niveaux de gris : SGLD (Spatial Gray Level Dependency) développé par Haralick [3]. C'est une méthode statistique qui consiste à construire des matrices de cooccurrence 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 [4].

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

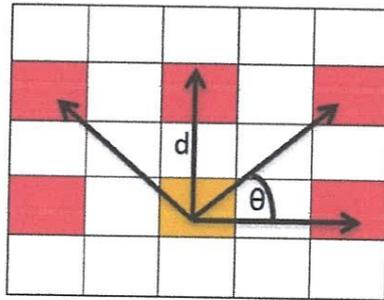


FIGURE 2.2 – La cooccurrence pour un pixel (jaune) avec $d = 3$ pixels et $\theta \in \{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}\}$.

Dans cette méthode, nous calculons la matrice de cooccurrence pour un ROI carré à huit angles différents : $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8} \text{ et } \frac{7\pi}{8}\}$ et à toutes les distances $d \in \{1, 2, 3, \dots, L/2\}$, où L est la longueur de ROI. Nous utilisons 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 cooccurrence sont générés pour chaque valeur de distance d , ces matrices sont additionnées pour obtenir une matrice de rotation invariant M_d :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad (2.5)$$

Ensuite, nous extrayons les huit des 14 caractéristiques de cooccurrence 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 cooccurrence matrice à la valeur de la distance d .

1. **Corrélation :**

$$x_1 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad (2.0)$$

2. **Homogénéité :**

$$x_2 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad (2.7)$$

3. **Variance :**

$$x_3 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu^d)^2 \quad (2.8)$$

4. **Contraste :**

$$x_4 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - j)^2 \quad (2.9)$$

5. Énergie :

$$x_5 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad (2.10)$$

6. Probabilité maximale :

$$x_6 = \max_{i,j} M_{i,j}^d \quad (2.11)$$

7. Somme moyenne :

$$x_7 = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i+1) \quad (2.12)$$

8. Cluster prominence :

$$x_8 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d + j - \mu_j^d)^4 \quad (2.13)$$

Si nous appliquons ces huit mesures pour chaque valeur de $\mathbf{d} \in [1, L/2]$, nous avons $(8 * L/2)$ caractéristiques. Ceci est très grandes quantités de données, dont la plupart sont redondants et non utile. Donc, nous utilisons les fonctions suivantes pour chaque fonctionnalité mesure de cooccurrences sur toutes les valeurs de \mathbf{d} , ces fonctions sont :

1. La moyenne :

$$x_{n+8} = \frac{1}{L \setminus 2} \sum_{d=1}^{L \setminus 2} x_n^d \quad (2.14)$$

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

2. Écart absolu moyen :

$$x_{n+16} = \frac{1}{L \setminus 2} \sum_{d=1}^{L \setminus 2} |x_n^d - x_{n+8}| \quad (2.15)$$

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

3. **Minimum :**

$$x_{n+24} = \text{Min}_{d=1}^{L \setminus 2} x_n^d \quad (2.16)$$

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

4. **Maximum :**

$$x_{n+32} = \text{Max}_{d=1}^{L \setminus 2} x_n^d \quad (2.17)$$

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

5. **Variance :**

$$x_{n+40} = \frac{1}{L \setminus 2} \sum_{d=1}^{L \setminus 2} (x_n^d - x_{n+8})^2 \quad (2.18)$$

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

2.2.1.3 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[5].

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

$$D = [\Delta x, \Delta y] = [x_a, y_b] - [x_n, y_m]. \quad (2.19)$$

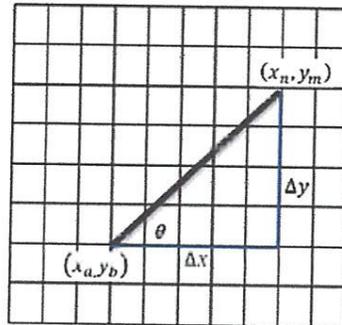


FIGURE 2.3 – Run Difference Matrix (RDM) is a function of r and level difference with the given direction θ .

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 :

$$RDM(r, gdif | \theta) = \# \{ ((x_a, y_b), (x_n, y_m)) : ((x_a, y_b), (x_n, y_m)) \in ROI, |G(x_n, y_m) - G(x_a, y_b)| = gdif \} / N$$

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

$$r = \sqrt{\Delta x^2 + \Delta y^2} \quad (2.20)$$

$$\theta = \tan^{-1} \left(\frac{\Delta y}{\Delta x} \right) \quad (2.21)$$

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

$$N = \# \{ ((x_a, y_b), (x_n, y_m)) : ((x_a, y_b), (x_n, y_m)) \in ROI \}$$

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

$$M_{RDM} = \sum_{\theta} RDM(r, gdif | \theta) \quad (2.22)$$

La matrice obtenue M_{RDM} n'est pas conçue pour l'extraction de caractéristiques mais nous pouvons 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 **DAD** (la distribution de distance moyenne) le vecteur :

$$DGD_{gdif} = \sum_{r=1}^{\frac{L}{2}} M_{RDM} \quad (2.23)$$

$$DOD = \sum_{gdf}^{G-1} M_{RDM} \cdot gdf \quad (2.24)$$

$$DAD_{gdf} = \sum_{r=1}^{\frac{L}{2}} M_{RDM} \cdot r \quad (2.25)$$

De ces trois caractéristiques, nous pouvons extraire autres caractéristiques tel que :

1. **Second Moment of DGD (SMG) :**

qui mesure la variation des différences de niveau de gris.

$$x_{59} = SMG = \sum_{gdf=0}^{G-1} (DGD(gdf))^2 \quad (2.26)$$

2. **Long distance emphasis for large difference (LDEL) :**

qui mesure l'importance des différences importantes sur une longue distance l'une de l'autre.

$$x_{61} = LDEL = \sum_{gdf=0}^{G-1} DAD(gdf)(gdf)^2 \quad (2.27)$$

2.2.1.4 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 Cooccurrence Matrix (GLCM) est une manière d'extraire les caractéristiques statistiques de texture du second degré [6]. 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 $P(i, j | d, \theta)$ est la fréquence relative avec laquelle deux pixels séparés 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 [7] :

1. 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 .
2. Ce nombre est inscrit dans la colonne i ème rangée et j du Pd matrice $[i, j]$.
3. 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 $[j, i]$.
4. Les éléments de Pd $[i, j]$ peuvent être normalisés en divisant chaque entrée par le nombre total de paires de pixels.
5. normalisé GLCM $N[i, j]$, définie par :

$$N[i, j] = \frac{P[i, j]}{\sum_i \sum_j P[i, j]} \quad (2.28)$$

2.2.1.5 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 [3] 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.

2.2.1.6 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 :

$$D = \log(N) / \log\left(\frac{1}{r}\right) \quad (2.29)$$

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 basé sur le **mouvement Brownien fractionnaire (fractional Brownian motion)** est souvent utilisé 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.

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 [8][9]. 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.

2.2.2 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, 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.

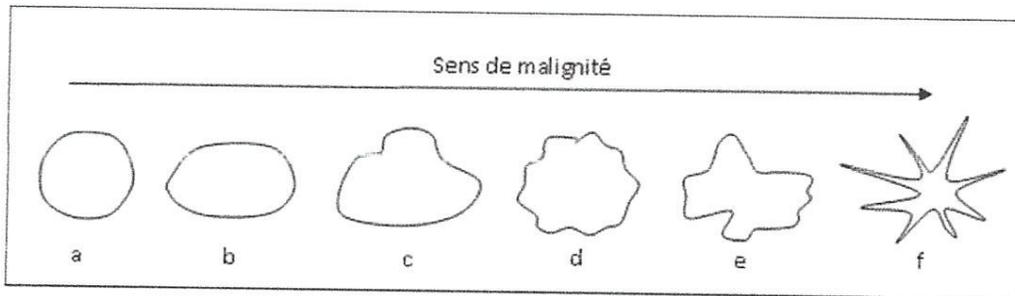


FIGURE 2.4 – Relation entre complexité du contour et malignité : a) masse circonscrite circulaire, b) circonscrite ovale, c) macro lobulée, d) micro lobulée, e) irrégulière, f) spéculée.

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 forme 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.

2.2.2.1 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

2.2.2.1.1 Le centre de gravité

Cette mesure est étroitement liée à 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 :

$$x_g = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.30)$$

$$y_g = \frac{1}{N} \sum_{i=1}^N y_i \quad (2.31)$$

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 :

$$x_g = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad (2.32)$$

$$y_g = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad (2.33)$$

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

2.2.2.1.2 L'aire

Parmi les descripteurs de forme les plus répandus, on cite l'aire de la masse, ce paramètre est calculé à partir du nombre de pixels contenus dans une lésion.

2.2.2.1.3 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 contour.

2.2.2.1.4 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 CEQ : 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 [10,11, 12, 13] :

$$C = \frac{\text{aire}(R \cap CEQ)}{\text{aire}(R)} \quad (2.34)$$

La figure explique davantage 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.

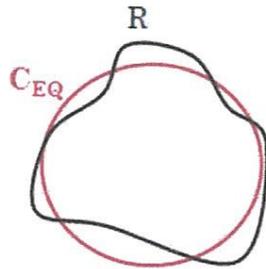


FIGURE 2.5 – Calcul de la mesure de circularité.

2.2.2.1.5 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 [11, 12,13]. On désigne par R la région étudiée et BE la boîte englobante qui est définie par le plus petit rectangle contenant la lésion étudiée. La mesure de rectangularité notée $Rect$ peut alors se définir comme suit :

$$Rect = \frac{aire(R)}{aire(Be)} \quad (2.35)$$

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.

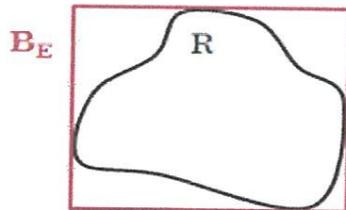


FIGURE 2.6 – Calcul de la mesure de rectangularité.

2.2.2.1.6 La compacité

Il s'agit d'une mesure de la complexité du contour vis à vis de l'aire [14]. Elle est notée par Com et est donnée par :

$$Com = \frac{p^2}{A} \quad (2.36)$$

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.

2.2.2.2 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.

2.2.2.2.1 La courbure

La notion de courbure (curvature en anglais) notée $Curv$ a été couramment utilisée Dans le cadre de l'analyse de formes dans plusieurs domaines. Elle a été reconnue pour sa capacité à caractériser les formes des objets. D'une manière générale, la courbure en un point donné A d'une courbe est définie en tant que l'inverse du rayon du cercle osculateur en A . Le cercle osculateur peut être obtenu ainsi : étant donné deux points B et C proches de A , on calcule le cercle unique passant par A , B et C . Dans le cas où ces points sont colinéaires, le cercle a un rayon infini et la courbure est alors nulle.

$$Curv = \frac{1}{r} \quad (2.37)$$

Le rayon du cercle osculateur est défini comme suit :

$$R = \frac{a, b, c}{\sqrt{(a+b+c)(a-b+c)(a+b-c)(b-a+c)}} \quad (2.38)$$

Avec $a = |AB|$, $b = |BC|$ and $c = |AC|$.

Récemment, plusieurs méthodes de calcul de la courbure ont été proposées. Nguyen et al [15] Ont amélioré l'estimation des cercles osculateurs proposée par [16] 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 [17] suggère de minimiser la courbure en respectant les contraintes Géométriques issues des directions tangentes calculées sur le contour. Malgouyres et al [18] ont suggéré d'appliquer une convolution binomiale afin d'obtenir un estimateur des cercles osculateurs convergent et adapté aux différents types d'images .

Kerautret et al [17] Proposent d'évaluer les trois méthodes décrites ci-dessus en s'appuyant sur différents tests de contours. Du point de vue temps d'exécution, la méthode proposée par [15] s'avère plus rapide que les autres. Du point de vue stabilité, la méthode proposée par [17] est plus robuste. Le descripteur de courbure tel qu'il est proposé par [17] est exploité dans le cadre d'un système DAOx afin d'être comparé a des descripteurs proposés dans la littérature.

2.2.2.2.2 Le nombre des protubérances et des dépressions importantes

Chen et al ont contribué avec 5 nouveaux descripteurs morphologiques visant à mettre en valeur la régularité du contour d'une masse et son allongement

Dans les images écho graphiques. Le premier descripteur proposé est intitulé le nombre des protubérances et des dépressions importantes (the number of substantial protuberances and dépressions (NSPD)). Comme son nom l'indique.

Ce descripteur consiste à compter le nombre des protubérances et des dépressions les plus importantes. Pour cela, on définit l'enveloppe convexe comme le plus petit ensemble de points convexes contenant la lésion.

Pour chaque point p_i du contour, on calcule le point d'angle polaire θ_i (Le point p_i est considéré comme un point convexe si le point d'angle polaire $\theta_i \geq \theta_p$ avec $\theta_p \in \{20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ\}$ un seuil positif prédéfini. L'idée de calculer les protubérances et les dépressions afin de déterminer le taux d'irrégularité d'une masse s'avère très intéressante. Etant données que les masses malignes sont plus spéculées et plus irrégulières, la valeur de NSPD correspondante sera plus élevée.

Toutefois, ce calcul dépend essentiellement du seuillage selon θ_i . Ainsi, plusieurs protubérances et dépressions qui sont graphiquement nettes ne sont pas nécessairement considérées.

2.2.2.2.3 Le squelette elliptique normalisé

ont utilisé la notion de squelette pour décrire les lésions mammaires dans les images écho graphiques. Le descripteur proposé, intitulé le Squelette elliptique normalisé connu sous l'acronyme anglais Elliptic Normalized Skeleton(ENS), est en effet le nombre de points dans le squelette. Ce nombre revient à la somme des points terminaux (PT), des points simples (PS) et des points multiples (PM) d'un squelette .

Cependant, cette entité n'est pas invariante aux transformations d'échelle de telle sorte que deux lésions identiques de tailles différentes ont différents nombres de points dans le squelette. Les auteurs ont alors suggéré de normaliser par le Périmètre de l'Ellipse Equivalente (noté PEEq). Le descripteur ENS est alors défini de la façon suivante :

$$ENS = \frac{Pt + Ps + Pm}{PEEq} \quad (2.39)$$

2.2.2.2.4 La longueur radiale normalisée

La longueur radiale est définie comme la distance euclidienne entre le centre de gravité de l'objet et l'ième pixel du contour.

Ainsi, la longueur radiale normalisée est le résultat de normalisation par le maximum de distance trouvée :

$$d_i = \frac{\sqrt{(x(i) - x_g)^2 + (y(i) - y_g)^2}}{\max(d(i))}, i = \{1, 2, \dots, N\} \quad (2.40)$$

avec $(x(i), y(i))$ et (x_g, y_g) les coordonnées du ième pixel et du centre de gravité respectivement. N est le périmètre de la masse.

Les auteurs ont développé un ensemble de six descripteurs basés sur la Longueur Radiale Normalisée connue sous la notation LRN (Normalized Radial Length en anglais.)

1. La moyenne de la longueur radiale normalisée (d_{avg}) :

La moyenne de la longueur radiale normalisée est en effet la mesure qui nous renseigne sur la façon dont le contour varie d'une manière macroscopique similaire à la mesure de circularité.

2. La déviation standard de la longueur radiale normalisée (σ) :

Cette mesure est un bon testeur d'irrégularité. En effet, plus le contour est irrégulier, plus la valeur de la déviation standard de la longueur radiale normalisée est élevée.

3. L'entropie (E) :

L'entropie est calculée à partir de l'histogramme de la longueur radiale. Le paramètre p_k est la probabilité que la LRN soit entre $d(i)$ et $d(i) + 1/N_{bins}$, avec N_{bins} le nombre de bins de l'histogramme normalisé, variant dans l'intervalle $[0,1]$ qui a été divisé en $N_{bins} = 100$. La mesure d'entropie calculée intègre simultanément la notion de circularité et d'irrégularité.

4. Le rapport de surface (A1) :

Le rapport de surface est une mesure du pourcentage de la partie de l'objet hors la région circulaire définie par la moyenne de la longueur radiale normalisée.

5. La rugosité (R) :

La rugosité est la mesure qui a pour objectif d'isoler la forme macroscopique de l'objet à partir de la structure fine des bords. Elle nous informe sur la moyenne entre les pixels voisins. Les contours irréguliers fournissent des valeurs élevées de rugosité.

6. Le taux de croisement en zéro (ZC1) :

(Zero Crossing Count) Ce taux est utilisé afin d'extraire les informations concernant les petites variations du contour. Il calcule le nombre de fois que la droite définie par la moyenne de LRN intercepte le contour de l'objet. C'est un indicateur du degré de spéculation du contour.

Les descripteurs LRN ont connu un très grand succès dans le domaine du diagnostic assisté par ordinateur. Ils ont fourni des résultats satisfaisants surtout avec des bords Ronds.

Finalement on présente dans le tableau les avantages et les inconvénients des descripteurs géométriques et spécifiques définis auparavant.

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

TABLE 2.1 – Les avantages et les inconvénients des descripteurs géométriques et spécifiques .

2.3 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. L'étude menée a montré que 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.

Ces descripteurs sont confrontés à certaines défaillances telles que la dépendance à la convexité de la forme, à certaines valeurs de seuil, à la valeur du plus grand et du plus petit lobe... Cette étude détaillée nous permet de tirer profit des points forts des descripteurs existants et de tenir compte des lacunes rencontrées.

Chapitre 3

Cancer du sein et pathologies mammaires

Sommaire

3.1	Introduction	56
3.2	Le cancer du sein	56
3.3	La Mammographie	57
3.4	Les pathologies mammaires	58
3.4.1	Les micros calcifications	58
3.4.2	Les masses	63
3.4.3	La distorsion architecturale	66
3.4.4	Les cas particuliers	66
3.4.5	Les signes associés	66
3.5	La classification des pathologies mammaires	66
3.5.1	La classification de Legal	66
3.5.2	La classification de Lanyi	68
3.5.3	La classification de Frouge	68
3.5.4	La classification de Moskowitz	69
3.5.5	La classification de Sigfusson	69
3.5.6	La classification BIRADS	70
3.6	Conclusion	73

Chapitre 3

CANCER DU SEIN ET PATHOLOGIES MAMMAIRES

3.1 Introduction

La mammographie est généralement la première procédure d'imagerie du sein utilisée dans l'évaluation des femmes qui présentent des signes ou symptômes mammaires qui peuvent être indicative de cancer.

Les résultats d'une mammographie doivent être communiqués au médecin traitant dans un format compréhensible cohérent qui comprend les résultats de l'imagerie, la probabilité de cancer, et de la recommandation pour le cours de l'action.

En effet, nous commençons, dans ce chapitre, par présenter la notion du cancer du Sein , puis Les outils de l'imagerie médicale utilisés pour le dépistage. Ensuite Les déférentes pathologies mammaires et La classification de ces pathologies mammaires.

3.2 Le cancer du sein

Les tumeurs malignes sont des cancers ,les cellules cancéreuses croissent et se divisent sans contrôles ,elles deviennent indifférenciés, ce qui signifie qu'elles perdent les caractéristiques morphologiques et fonctionnelles du tissu d'origine. Elles peuvent envahir et détruire les tissus voisins et les organes. De plus, elles peuvent se mobiliser et pénétrer dans le sang ou le réseau lymphatique [19].

Le cancer du sein est une tumeur maligne de la glande mammaire. Autrement dit, c'est un cancer qui naît dans les unités cellulaires dont la fonction est de sécréter le lait, les unités ducto-lobulaires du sein, essentiellement chez la femme (le cancer du sein survient 200 fois moins souvent chez l'homme, qui possède lui aussi des seins, bien qu'atrophies).

C'est le cancer le plus fréquent chez la femme, représente l'une des premières causes de décès parmi les cancers gynécologiques. Une forte majorité des cancers prennent naissance dans le tissu glandulaire et pas dans n'importe quelle région du sein. En effet, 50 % des cancers du sein se situent dans le (QSE) de la glande, 15 % dans le (QSI), 11 % dans le (QIE), 18 % au niveau du mamelon et à peine 6 % dans le (QII).

Le nombre de personnes atteintes a progressé légèrement, mais régulièrement, au cours des 3 dernières décennies. Par contre, le taux de mortalité a continuellement diminué au cours de la même période, grâce aux progrès réalisés en matière de dépistage, de diagnostic et de traitement [A].

3.3 La 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.

La mammographie est le meilleur moyen de dépistage du cancer du sein. Elle permet de révéler des petites tumeurs qui ne sont pas toujours détectables par la seule palpation des seins. Elle est recommandée tous les deux ans pour les femmes de plus de 50 ans. Pour les femmes de moins de 50 ans, elle peut être recommandée dans certaines conditions. Demandez conseil à votre médecin. La mammographie n'est pas un examen infaillible mais elle permet tout de même de dépister 90% des lésions du cancer du sein. La mammographie détecte également des lésions autres que celles des cancers. C'est pour cela qu'il est nécessaire de passer d'autres examens pour s'assurer qu'il s'agit vraiment d'un cancer du sein .

L'appareil dédié à la réalisation d'une mammographie est le mamographe , Les composants d'un mamographe est illustré dans la figure

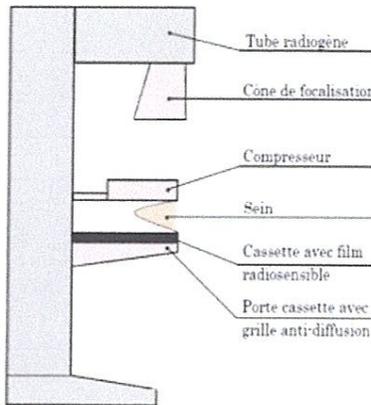


FIGURE 3.1 – Les composants d'un mammographe.

3.4 Les pathologies mammaires

3.4.1 Les micros calcifications

Une micro calcification est un dépôt de sels de calcium composé des substances chimiques $\text{Ca}_3(\text{PO}_4)_2$, CaCO_3 et $\text{Mg}_3(\text{PO}_4)_2$.

Ces substances sont très radio-opaques et se traduisent, dans les clichés mammographiques, par de petits points clairs. Les caractéristiques qui distinguent les micros calcifications des autres éléments sont leur fort contraste et leur petite taille ($< 0,5$ mm). Une fois leur taille dépasse 1 mm, on les appelle des macros calcifications et elles sont souvent bénignes.

Les micros calcifications 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. La description des micros calcifications permettant de décider de leur degré de suspicion inclut simultanément le critère de morphologie, de distribution et de nombre.

3.4.1.1 Morphologie

L'analyse de la morphologie est très importante [20][21]. Elle permet le plus souvent de séparer les micros calcifications bénignes et malignes... On détaille dans ce qui suit les différents types de micros calcifications et des exemples explicatifs de chaque cas dans la figure.

– **Micro calcifications cutanées ou dermique :**

elles présentent typiquement un centre clair. Des clichés en incidence tangentielle sont souvent utilisés pour confirmer la localisation cutanée de ces micros calcifications.

– **Micro calcifications vasculaires :**

ces micros calcifications en rails ou linéaires sont associées à des structures tubulaires.

– **Micro calcifications grossières ou coralliformes :**

elles sont de grande taille (> a 2-3 mm de diamètre).

– **Micro calcifications en bâtonnets :**

associées a une ectasie canal aïre et sont alors dirigées vers le mamelon. Elles mesurent habituellement plus de 1 mm de large et peuvent présenter un centre clair si le dépôt calcique se fait dans la paroi du canal.

– **Micro calcifications rondes :**

elles ont une forme ronde et peuvent être de tailles variables. Lorsqu'elles mesurent < de 0.5mm, elles sont dites punctiformes ou pulvérulentes.

– **Micro calcifications a centres clairs :**

leur taille peut s'étendre de 1 mm a plus de 1 cm. Elles sont rondes ou ovales, à surface lisse et à centre clair.

– **Micro calcifications en coquille d'œuf ou pariétales :**

Très fines apparaissent comme des dépôts calciques sur la surface d'une

sphère. Vu dans l'axe du rayonnement X, ce dépôt mesure moins de 1 mm.

– **Micro calcifications a type de lait calcique :**

elles sont sédimentées dans le fond de kystes. En utilisant l'incidence cranio-caudale, elles sont souvent difficiles a discerner. Par contre, l'incidence de profil permet de démontrer leurs formes caractéristiques : semi-lunaires, en croissants, curvilignes ou linéaires.

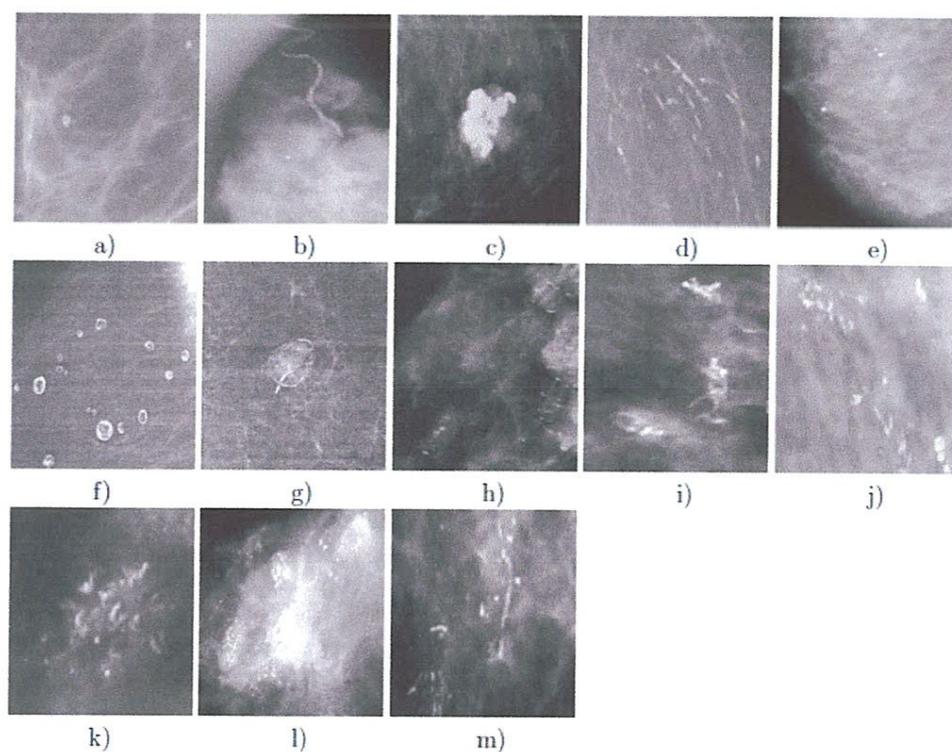


FIGURE 3.2 – Les différents types de microcalcifications : a) cutanées ou dermiques, b) vasculaires, c) grossières ou coralliformes, d) en bâtonnets, e) rondes, f) a centres clairs, g) en coquille d'oeuf ou pariétales, h) a type de lait calcique, i) de suture, j) dystrophiques, k) amorphes ou indistinctes, l) fines polymorphes, m) linéaires ramifiées .

– **Micro calcifications de suture :**

elles correspondent a des dépôts calciques sur du matériel de suture, sont typiquement linéaires ou tubulaires et présentent fréquemment des nœuds.

– **Micro calcifications dystrophiques :**

elles mesurent habituellement plus de 0.5 mm de diamètre et sont de formes irrégulières. Elles présentent parfois un centre clair.

– **Micro calcifications amorphes ou indistinctes :**

elles sont souvent plus ou moins rondes ou en forme de flocons. Elles sont de petites tailles et généralement a contours vagues sans forme spécifique.

– **Micro calcifications fines et polymorphes :**

elles sont habituellement mieux visibles que les micros calcifications amorphes. Elles sont irrégulières de taille et de forme variables mesurant moins de 0.5 mm de diamètre.

– **Micro calcifications linéaires et ramifiées :**

mesurent moins de 0.5 mm d'épaisseur. Elles sont irrégulières et de formes parfois linéaires ou curvilignes.

3.4.1.2 Distribution

Est un critère fondamental. Elle présente leur répartition dans le sein et joue un rôle important dans la prise de décision de la malignité [21].

Les différentes distributions possibles des micros calcifications sont détaillées dans ce qui suit.

– **Micro calcifications diffuses/éparses :**

les micros calcifications sont distribuées de façon aléatoire dans l'ensemble du sein.

– **Distribution régionale :**

dispersées dans un large volume du tissu mammaire et ne présentent pas une distribution canal aire .

– **Micro calcifications groupées, en amas ou en cluster :**

multiples micros calcifications (au moins cinq) occupent un petit volume tissulaire.

– **Distribution linéaire :**

disposées les unes derrière les autres sous forme d'une ligne.

– **Distribution segmentaire :**

elle suggère des dépôts calciques dans des canaux galactophores ainsi que leurs branches ce qui évoque la possibilité d'un cancer mammaire étendu.

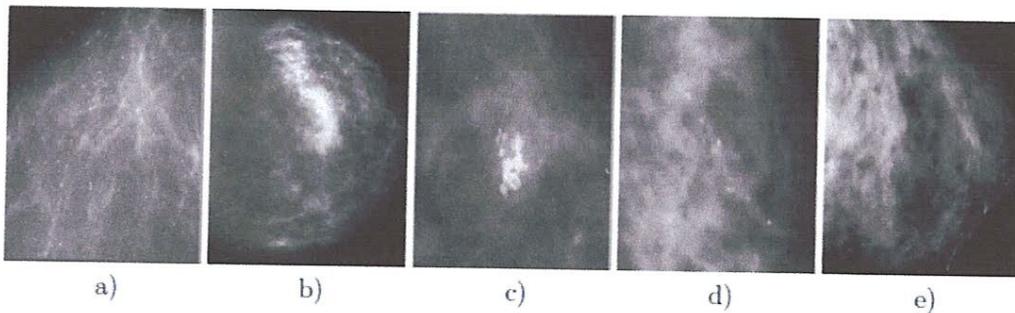


FIGURE 3.3 – Les différentes distributions des microcalcifications : a) Mcs diffuses ou éparses, Répartition : b) Régionale, c) En cluster, d) Linéaire et e) Segmentaire.

3.4.2 Les masses

Une opacité ou une masse est une lésion importante occupant un espace, vue sur deux incidences différentes. Elle est caractérisée par :

3.4.2.1 La forme

Selon la description du BIRADS [22], les masses mammaires peuvent avoir la forme ronde (figure a), ovale (figure b), lobulée (figure c) ou irrégulière (figure d).

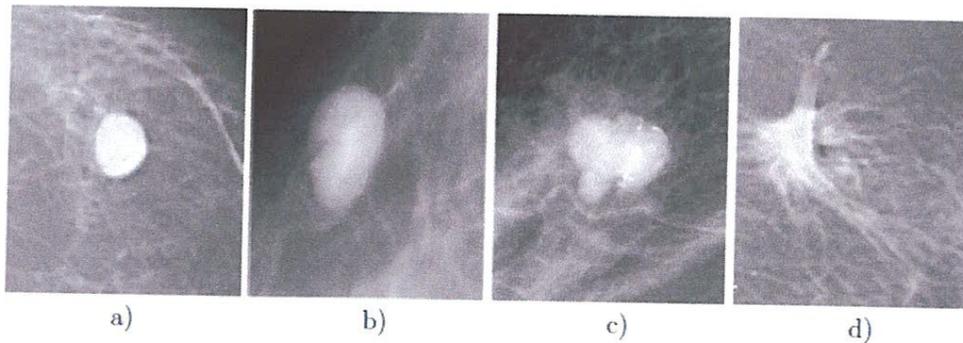


FIGURE 3.4 – Les différentes formes possibles d’une masse : a) Ronde, b) Ovale, c) Lobulée , d) Irrégulière .

– **Ronde :**

Il s’agit de masse sphérique, circulaire ou globuleuse.

– **Ovale :**

Elle présente une forme elliptique (ou en forme d’œuf).

– **Lobulée :**

La forme de la masse présente une légère ondulation.

– **Irrégulière :**

Cette appellation est réservée aux masses dont la forme est aléatoire

3.4.2.2 Le contour

Le contour des masses mammaires est soit circonscrit (figure a), soit micro lobulé (Figure b), soit masqué (figure c) soit indistinct (figure d), soit spiculé (figure e).

– **Circonscrit :**

Il s'agit d'une transition brusque entre la lésion et le tissu environnant. Le contour est alors net et bien défini.

– **Micro lobulé :**

Dans ce cas, de courtes dentelures du contour créent de petites ondulations.

– **Masqué :**

Un contour masqué est un contour qui est caché par le tissu normal adjacent.

– **Indistinct :**

Dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre a une infiltration.

– **Spiculé :**

La masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse (les spicules).

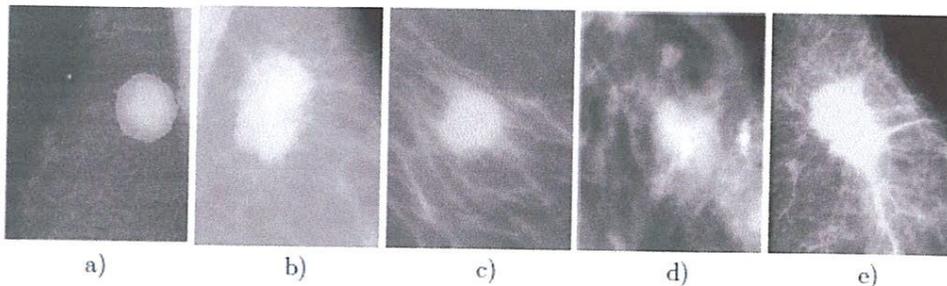


FIGURE 3.5 – Les différents contours possibles d'une masse : a) Circonscrit, b) Microlobulé, c) Masqué, d) Indistinct et e) Spiculé .

3.4.2.3 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 [23]. La classification BIRADS de l'ACR définit 4 classes de la composition du sein.

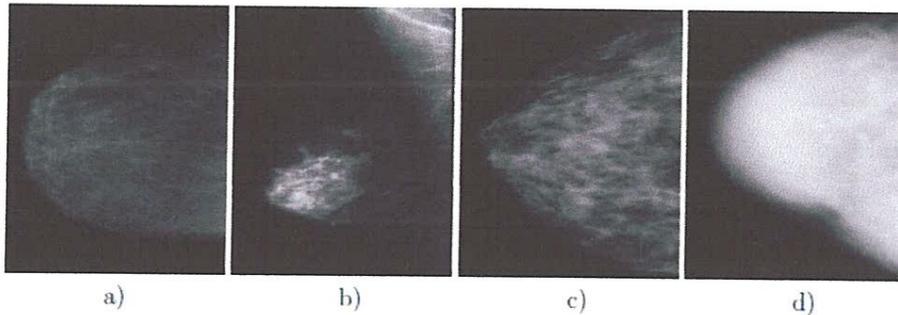


FIGURE 3.6 – – Densité mammaire selon le lexique BIRADS : a) Densité < 25 %, b) Densité entre 25 et 50%, c) Densité entre 50 et 75% et d) Densité > 75%.

– **Stade 1 :**

Le sein est presque entièrement graisseux et homogène, radio transparent et facile à lire (moins de 25 % de la glande mammaire).

– **Stade 2 :**

Il y a des opacités fibro glandulaires dispersées. Le sein est graisseux et hétérogène (approximativement 25 à 50 % de la glande mammaire).

– **Stade 3 :**

Le tissu mammaire est dense et hétérogène (approximativement 51 à 75 % de la glande mammaire).

– **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).

3.4.3 La distorsion architecturale

Il s'agit d'une rupture de l'architecture normale sans masse patente. Le BI-RADS donne trois exemples : fines lignes ou spécules irradiant à partir d'un point, rétraction locale, ou distorsion d'un bord du parenchyme.

3.4.4 Les cas particuliers

- Structure tubulaire asymétrique ou galactophore dilaté isolé.
- Ganglion intra mammaire.
- Asymétrie de répartition du parenchyme.
- Asymétrie focale de densité.

3.4.5 Les signes associés

Ils peuvent être associés aux masses, aux calcifications : rétraction cutanée, mamelon aigre, épaissement cutané, local ou diffus, lésion cutanée ou adénopathie dense.

3.5 La classification des pathologies mammaires

Il existe plusieurs classifications de ces lésions selon L'étude de la morphologie à savoir la classification de Le Gal [23], de Lanyi [20]...et de BIRADS [21][24] .

3.5.1 La classification de Legal

En 1976, Le Gal du Service de Radiodiagnostic de l'institut Curie a Paris, a conçu la classification dite de Le Gal [23]. Elle décrit cinq types morphologiques qui ont une valeur prédictive de malignité croissante :

– **Type 1 :**

Mcs annulaires rondes à centre clair radio-transparentes, arciformes ou polyédriques. Risque de cancer du sein quasi nul.

– **Type 2 :**

Mcs punctiformes, rondes et de tailles variables. Risque de carcinome : 22%..

– **Type 3 :**

Mcs poussiéreuses, trop fines pour préciser leur forme elles donnent une image de semis de poudre calcaire, pulvérulentes. Risque de cancer : 36%..

– **Type 4 :**

Mcs irrégulières aux contours anguleux, différents d'une mcs à l'autre associées à un risque de cancer : 56% .

– **Type 5 :**

Mcs vermiculaires ou branchées elles ont la forme d'un bâtonnet souvent irrégulier. Risque de carcinome : 90% . .

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

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

3.5.2 La classification de Lanyi

Lanyi 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 .

3.5.3 La classification de Frouge

Quatre groupes de microcalcifications évocatrices de carcinomes canaux in situ sont individualisés. Ils tiennent compte de la morphologie et de la répartition dans le sein [C] .

– **Le type A :**

est caractérisé par des calcifications de type 2 et 3 de Legal en concentration importante mais regroupées au sein d'un foyer unique comportant de 4 à 20 éléments. C'est celui qui pose le plus de problèmes diagnostiques.

– **Le type B :**

correspond à la présence de plusieurs foyers de type A dans un même quadrant témoignant d'une propagation à plusieurs lobules dans un même arbre galactophorique.

– **Le type C :**

correspond à l'aspect habituel du comédocarcinome avec des calcifications de type 4 et 5 regroupées en un foyer unique et avec une concentration importante.

– **Le type D :**

d'interprétation difficile, correspond à des calcifications diffuses, de topographie segmentaire peu caractéristique. Les calcifications sont de type 1 et 2 de Le Gal et de concentration faible.

3.5.4 La classification de Moskowitz

La classification de Moskowitz permet de classifier Les microcalcifications en 4 catégories (voir le tableau) [D] .

Aucun	- Petites, irrégulières, semi-lunaires. - Suivi régulier.
Minime	- Nombreuses, irrégulières, en amas. - Biopsie.
Modéré	- Punctiformes, incurvées, en zone large. - Biopsie
Elevé	- Linéaires, peu denses. - Biopsie.

TABLE 3.1 La classification de Moskowitz.

3.5.5 La classification de Sigfusson

Cette classification permet de classifier Les microcalcifications en 4 catégories (voir le tableau)[D].

Aucun	- Circulaires, moulées, bénignes mais 6 % de CLIS.
Minime	- Irrégulières. - (18 % de cancer du sein).

Modéré	- Parfois irrégulières ou linéaires. - (37 % de cancer du sein).
Elevé	- Très irrégulières, linéaires, branchées. - (96 % de cancer du sein).

TABLE 3.2 – La classification de Sigfusson.

3.5.6 La classification BIRADS

Pour palier les inconvénients de la classification de Le Gal. En novembre 1998 l'American College of Radiology (ACR) a établi le système Américain BIRADS (Breast Imaging Reporting and Data System) qui a été rédigé par un groupe d'experts réunis par l'ANAES. Pour L'objectif de :

- standardiser les comptes rendus mammographique.
- faciliter la comparaison entre les examens successifs, venant ou non de sources différentes .
- recueillir les données pour suivre les pathologies, évaluer les pratiques et les résultats du dépistage à plus grande échelle.

Ce système permet de classier l'image mammographique en six catégories en fonction du degré de suspicion de leur caractère pathologique (voir le tableau).

ACR 0		-Cliché insuffisant pour vérifier la présence ou non de lésions.
ACR 1	Masses	- Opacité ronde avec macrocalcifications.
	Mcs	- Absence de microcalcifications suspectes.
ACR 2	Masses	- Absence de masses suspectes. - Ganglion intra mammaire. - Images de densité graisseuse ou mixte. - Opacités rondes correspondant a un kyste en échographie.
	Mcs	- Macrocalcifications cutanées. - Calcifications punctiformes régulières diffuses. - Calcifications sur matériel de suture. - Macrocalcifications sans opacité. - Mcs annulaires, semi-lunaires ou sédimentées.
ACR 3	Masses	- Masse circonscrite ronde ou ovale sans microlobulations non liquidienne en échographie.
	Mcs	- Mcs rondes ou punctiformes régulières ou pulvérulentes, peu nombreuses, en petit amas rond isolé. - Petit amas rond ou ovale de calcifications amorphes et peu nombreuses.
ACR 4	Masse	- Masse spiculée sans centre dense. - Masse ronde ou ovale a contour lobulé ou masqué.
	Mcs	- Mcs pulvérulentes groupées et nombreuses. - Mcs irrégulières, polymorphes ou granulaires, peu nombreuses.

ACR 5	Masses	<ul style="list-style-type: none"> - Masse spiculée a centre dense. - Masse a contour flou et irrégulier.
	Mcs	<ul style="list-style-type: none"> - Mcs associées a une anomalie architecturale ou a une opacité. - Groupement de Mcs quelle que soit leur morphologie, dont la topographie est galactophorique. - Mcs vermiculaires, arborescentes ou irrégulières polymorphes ou granulaires nombreuses et groupées. - Mcs groupées ayant augmenté en nombre ou Mcs dont la morphologie et la distribution sont devenues plus suspectes.

TABLE 3.3 – Classification BIRADS de l'ACR .

ACR 0	Nécessité d'investigations complémentaires : telles que cliché avec compression centrée, agrandissement, incidence particulière échographie.
ACR 1	Négative : mammographie normale, aucune masse ou calcification suspecte n'est présente.
ACR 2	Lésions bénignes : ne nécessitant ni surveillance ni examen complémentaire
ACR 3	Lésions probablement bénignes : une surveillance clinique et radiologique a court terme est conseillée.
ACR 4	Anomalies suspectes : une biopsie devrait être envisagée. Ces lésions n'ont pas un aspect typique de cancer, mais peuvent néanmoins correspondre a une lésion maligne.

ACR 5	Lésions fort suspectes de malignité : l'anomalie est évocatrice d'un cancer.
--------------	---

TABLE 3.4 – Conduite a tenir pour chaque classe de l'ACR.

La classification de l'ACR résume les formes des différentes masses et des différentes micros calcifications, leur texture, les différents aspects de la distorsion architecturale ainsi que leur degré de malignité. Une fois que le radiologue arrive à reconnaître la catégorie d'une mammographie, il sait automatiquement les directives et les recommandations associées à cette classe.

3.6 Conclusion

Dans ce chapitre nous avons présenté, la notion du cancer du sein et la mammographie .

Une grande attention a été consacrée à l'étude des pathologies mammaires à savoir les masses et les micros calcifications. Une telle étude est fortement associée a la présentation des standards adoptés par les radiologues pour classifier les lésions mammaires en bénignes /malignes.

Chapitre 4

Conception et Application

Sommaire

4.1	Introduction	75
4.2	Environnement et outils de développement	75
4.2.1	Plateformes utilisées	75
4.2.2	Langage et environnement	75
4.2.3	Choix du SGBD	77
4.2.4	Java PersistenceAPI (JPA)	77
4.3	Description générale du système	78
4.3.1	Gestion de la base DDSM	80
4.3.2	Calcul des caractéristiques	84
4.3.3	Développement d'un RNA	87
4.4	Réalisation et interface	89
4.5	Résultats et discussion	93
4.6	Conclusion	95

Chapitre 4

CONCEPTION ET APPLICATION

4.1 Introduction

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

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 7. Les clichés mammographiques sont décompressés sous Linux elementary OS.

4.2.2 Langage et environnement

Langage de programmation JAVA avec l'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 fonctionnent pas si Java n'est pas installé et leur

nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable. [B]

La version d'Eclipse utilisée est « Luna 4.4 » sortie de 25 juin 2014 et est composée de 76 projets cette dernière apporte quelques nouveautés majeur :

- Le support natif de Java 8.
 - L'amélioration du workbench : un nouveau thème sombre, la possibilité de séparer l'éditeur d'un même fichier en deux parties (horizontalement ou verticalement), l'affichage par défaut du nombre de lignes. . .).
 - L'amélioration du plugin EGit
 - L'intégration d'un terminal (TCF Terminal) permettant d'utiliser le terminal du système, de se connecter en SSH ou Telnet.
 - Le projet Sirius, qui permet de créer des outils de modélisation sur mesure.
 - Une amélioration de la bibliothèque EGIT, l'implémentation Java de GIT.
 - L'outil graphique de fusion a également été amélioré
- La constatation d'un de ces signes impose une consultation médicale sans retard. Le médecin jugera des examens complémentaires nécessaires.



FIGURE 4.1 – Démarrage d'Eclipse.

4.2.3 Choix du SGBD

Nous avons choisis le système de gestion des bases de données Derby. Derby est un système de bases de données relationnelles de la fondation Apache sous licence Open Source dont l'origine était un projet d'IBM nommé Cloudscape. Ce SGBD est également fourni par SUN dans son kit de développement Java (depuis le JDK 6) mais également avec NetBeans et GlassFish. Dans ce cas, il se nomme JavaDB.

Derby a la particularité principale d'avoir été écrit entièrement en Java tout en fournissant les fonctionnalités d'un système de gestion de bases de données relationnelle complet avec implémentation du langage SQL, gestion des transactions, triggers et procédures stockées.

4.2.4 Java PersistenceAPI (JPA)

JPA pour Java Persistence API est une API qui a pour but de résoudre le problème de persistance d'objets dans une base de données relationnelle. Ce mécanisme qui gère le lien entre des objets d'une application avec les tables de la base de données se nomme **Object-RelationalMapping** ou **ORM**.

Les classes dont l'état doit être sauvegardé sont dites entités et sont marquées dans le code source par l'annotation **@Entity**. Les fonctions permettant de gérer la persistance et de faire le lien avec la base de données physique se trouvent implantées par une classe particulière qui se nomme **EntityManager**.

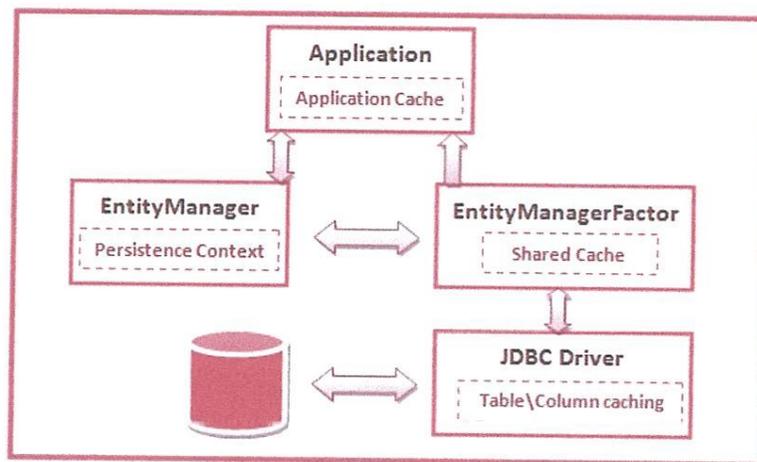


FIGURE 4.2 – Une application qui utiliserait une classe Client persistante.

4.3 Description générale du système

Notre système se compose principalement de trois étapes, l'objectif de la première étape est de créer une base de données relationnelle basée sur la technologie JPA (Java Persistence API) pour la gestion de la base mammographique DDSM, cette phase facilite la tâche d'extraction d'informations relatifs aux clichés mammographiques et aux anomalies.

dans la deuxième phase un ensemble de caractéristiques sont calculées à travers des méthodes connues présentés dans le chapitre 2 (SGLDM et RDM) pour renforcer leurs pouvoir en décrire les caractéristiques pertinentes des masses mammaires.

Et la troisième étape consiste à créer un réseau de neurones artificiels adéquat au problème traité afin de tester les performances de notre système.

La conception générale de notre application se résume dans le schéma suivant :

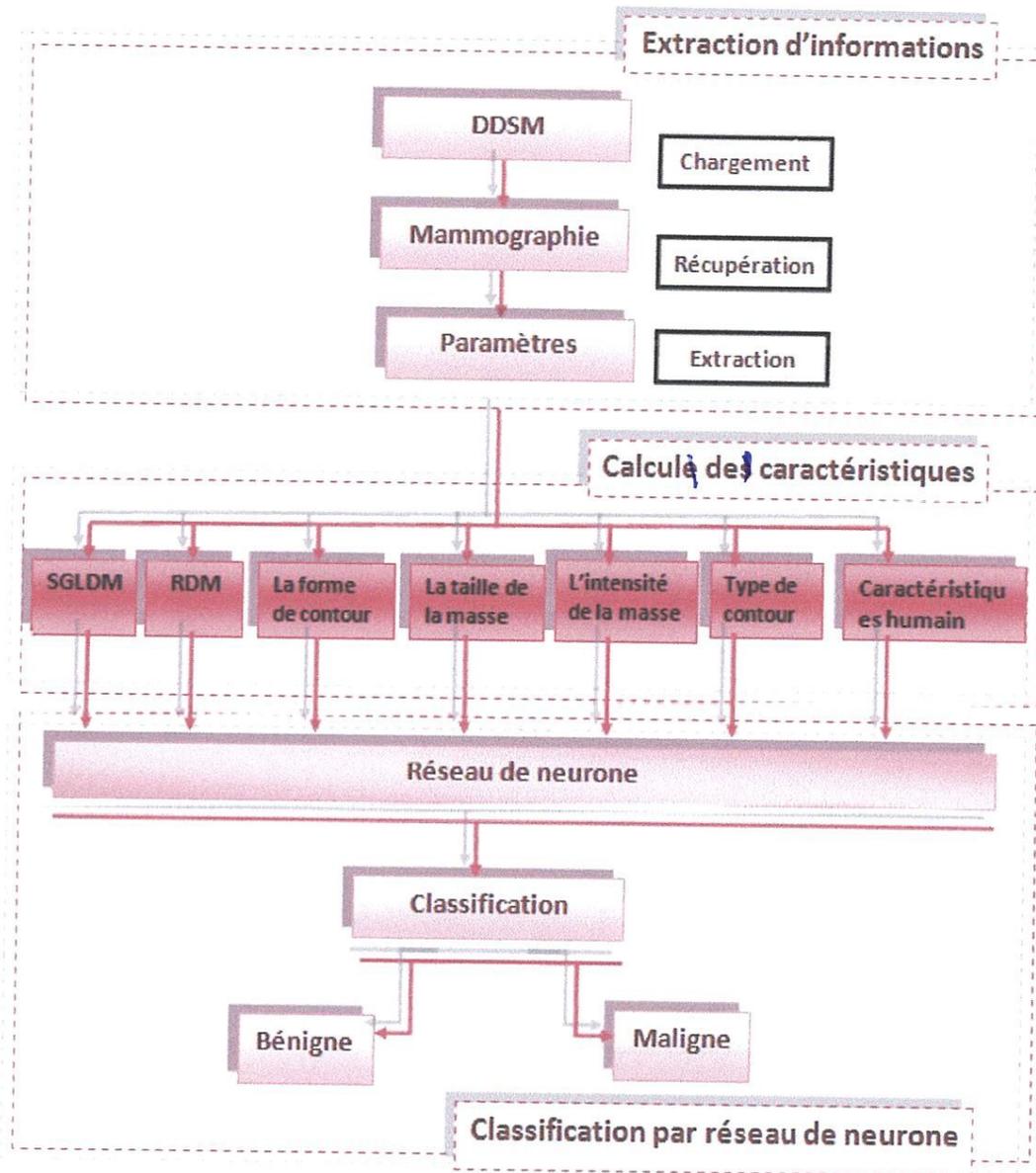


FIGURE 4.3 – Schéma générale de notre approche.

4.3.1 Gestion de la base DDSM

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 on 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 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 de 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 .

Dans la figure qui suit les informations qui sont cadrés.

ics_version	1.0								
filename	B-3024-1								
DATE_OF_STUDY	2 7 1995								
PATIENT_AGE	42								
FILM									
FILM_TYPE	REGULAR								
DENSITY									
DATE_DIGITIZED	7 22 1997								
DIGITIZER	LUMISYS SELECTED								
LEFT_CC_LINES	4696	PIXELS_PER_LINE	3024	BITS_PER_PIXEL	12	RESOLUTION	50	NON-OVERLAY	
LEFT_MLO_LINES	688	PIXELS_PER_LINE	3048	BITS_PER_PIXEL	12	RESOLUTION	50	OVERLAY	
RIGHT_CC_LINES	624	PIXELS_PER_LINE	3056	BITS_PER_PIXEL	12	RESOLUTION	50	NON-OVERLAY	
RIGHT_MLO_LINES	664	PIXELS_PER_LINE	3120	BITS_PER_PIXEL	12	RESOLUTION	50	OVERLAY	

FIGURE 4.4 – Les informations importantes dans un fichier ics.

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,

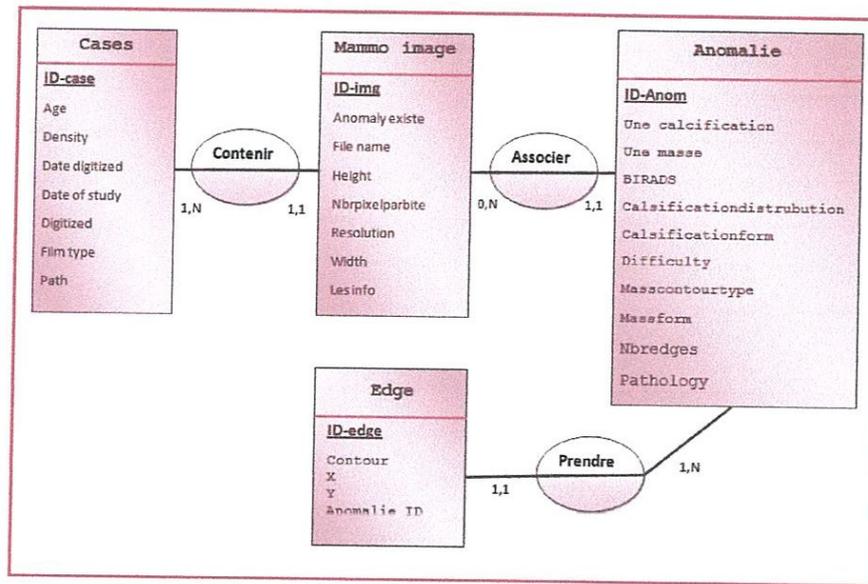


FIGURE 4.6 – Le modèle conceptuel des données.

4.3.1.4 Conversion du LJPEG au LJPEG1

LJPEG nommé aussi JPEG-LS (souvent surnommé Lossless JPF/G) 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'en-tête. Les dimensions de l'image (hauteur et largeur) sont précisées dans le fichier "ics" pour ce cas et cette décompression est faite sous Linux et l'instruction suivante sert à décompresser une seule image.

L'image est stockée étendue en tant que données binaires brutes dans un fichier qui ne contient pas d'informations d'en-tête. Les dimensions de

l'image (hauteur et largeur) sont précisées dans le fichier "ics" pour ce cas, et cette décompression est faite sous L'Unix et L'instruction suivante sert à décompresser une seul image.

USAGE : jpeg -d -s 'image file name'

EXAMPLE : jpeg -d -s A_ 0782.RIGHT_MLO .LJPEG

Malheureusement cette commande permet de décompresser qu'une image à la fois, pour cela nous avons transcrit un bash sous linux afin de décompresser toutes les imagesmammographique d'un seul coup.

```
#!/bin/bash
echo
echo "Processing directory $1"
echo
for i in $(find . -name "*.LJPEG"); do
echo "> Processing file : $i"
./jpeg -d -s $i &> /dev/null
done
echo
echo done.
```

FIGURE 4.7 – Le code de décompression d'images Ljpeg.

4.3.2 Calcule des caractéristiques

Les caractéristiques choisis pour le procédé de classification devraient refléter la spécificité des masses apparaissant dans la mammographie aussi précise que possible. Le choix des paramètres est important pour effectuer une classification efficace. Il existe plusieurs paramètres pour décrire les masses.

4.3.2.1 Les caractéristique RDM (RunDifference Matrix)

Un pseudo code illustrant les caractéristiques calculé par la méthode RDM présenté dans le chapitre 2 qui sont :

- **DAD** (Distribution of Average Distance).
- **DGD** (Distribution of Gray level Difference).
- **SMG**(Second Moment of DGD).
- **LDEL** (Long Distance Emphasis for large Difference).

Entrée :

- La direction utilisée pour la cooccurrence $\theta = \{0^\circ, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$
- ROI (region of interest)

Les caractéristiques pour être calculé $k = \{DGD, DAD, SMG, LDEL\}$.

Debut

Pour chaque angle θ faire

Pour tout les distance $d \in \{1, 2, 3, \dots, L/2\}$

Pour $i = 1$ jusqu'à $\text{length}(k)$ faire

Calculer $(k(i))$;

Fin ;

FIGURE 4.8 – Un pseudo code de la méthode RDM.

4.3.2.2 Les caractéristiques SGLDM (Spatial Gray Level Dependence Method)

Cette méthode utilisée pour calculer les caractéristiques désirées qui sont : Variance, Maximum, Minimum, Mean, Mean absolute deviation. Un pseudo code illustrant cette technique .

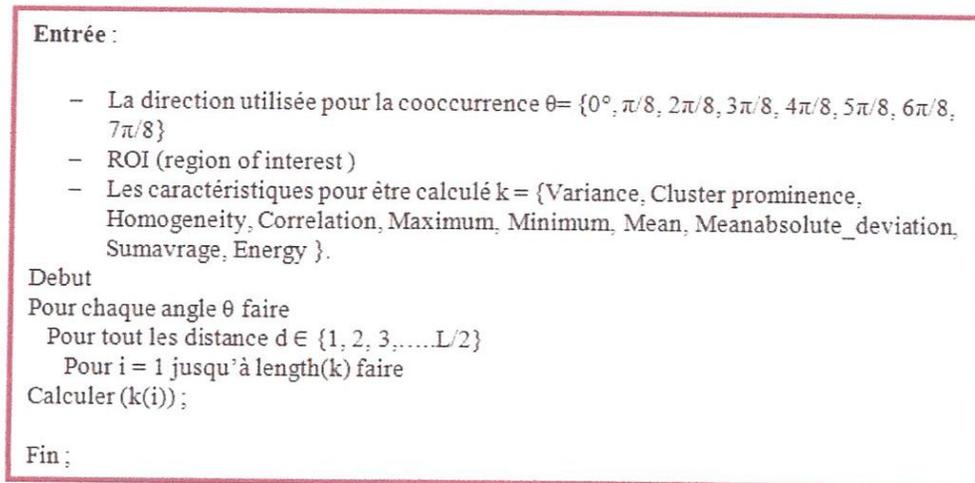


FIGURE 4.9 – Un pseudo code de la méthode SGLDM.

4.3.2.3 Les Caractéristiques humain

L'âge et densité sont des caractéristiques humaines, qui sont donnés par DDSM [44]. La description de l'interprétation de ces traits humains est basée sur Breast Imaging Reporting and Data System (BI-RADS) [45], qui sont utilisés comme un guide pour la normalisation des rapports mammographiques

Valeur de densité	Interprétation BI-RADS
1	Le sein est presque entièrement de matières grasses.
2	Il ya 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 4.1 – Interprétation BIRADS pour les valeurs de densité du sein .

4.3.2.4 Autres caractéristiques

- **L'intensité de la masse :**

Est égale à la moyenne d'intensités des pixels de la zone englobée par le contour .

- **Type de contours de la masse :**

le type de contours aussi lié à l'énergie externe du contour . Car l'énergie externe prend une valeur importante lorsque la masse a des contours bien définis et circonscrits, par contre on obtient une faible énergie externe dans le cas des contours mal définis ou effacés.

- **La taille de la masse :**

La taille de la masse est représentée par la zone englobée par le contour.

- **La forme du contour :**

la forme de la masse est fortement liée à la courbure et la continuité du contour . C'est-à-dire l'énergie interne La courbure qui a un comportement comme une plaque mince. Il prend une valeur importante lorsque la courbe s'incurve rapidement c'est-à-dire pour l'obtention de coins et les formes complexes, par contre lorsque cette énergie est faible la courbe tendra vers un cercle dans le cas d'un contour fermée .La formule de la courbure est la suivante :

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

4.3.3 Développement d'un RNA

Le vecteur des caractéristiques pour chaque image de base de données a été calculé précédemment. Il apparaît indispensable à ce stade de développer le RNA. Le développement du RNA de ce travail va entièrement s'effectuer à de la bibliothèque Neuroph. Ce Framework peut être utilisé pour créer et

former des réseaux de neurones dans des programmes Java de manière intuitive. Il en est de même pour l'entraînement et le test de celui-ci.

4.3.3.1 Architecture générale du RNA

Cette étape se base sur une architecture classique perceptron multi couches (PMC). Cette architecture est particulièrement bien adaptée à la classification d'ensembles de vecteurs d'entrée (vecteur des caractéristiques) en ensembles de vecteurs de sortie correspondants (bénigne ou maligne).

Dans sa conception standard, l'architecture multi-layer perceptron comporte trois couches de neurones. Une couche d'entrée, une couche cachée et une couche de sortie. Pour tirer profit au maximum du RNA, il est primordial de définir intelligemment le nombre de neurones et leurs fonctions dans chacune des couches créées.

4.3.3.2 La couche de sortie

Le RNA devra être en mesure de déterminer si la masse est maligne ou bénigne. La couche de sortie devra donc avoir les spécificités suivantes :

- 01 neurone pour l'identification des masses malignes ;
- 01 neurone pour l'identification des masses bénigne ;

4.3.3.3 La couche d'entrée

La couche d'entrée doit comporter les neurones recevant les caractéristiques calculés de la masse. Ces entrées seront transmises au réseau par l'intermédiaire des fichiers d'entraînement et de test.

Les entrées choisies sont les 17 caractéristiques présentées dans la section précédente.

4.3.3.4 La couche cachée

La topologie PMC comporte encore une ou plusieurs couches cachées. Le nombre de neurones ainsi que le nombre de couches est un choix arbitraire dépendant des connaissances de l'utilisateur. Une seule couche cachée comportant 17 neurones semble être un choix raisonnable.

4.4 Réalisation et interface

L'interface de notre application est comme suit :

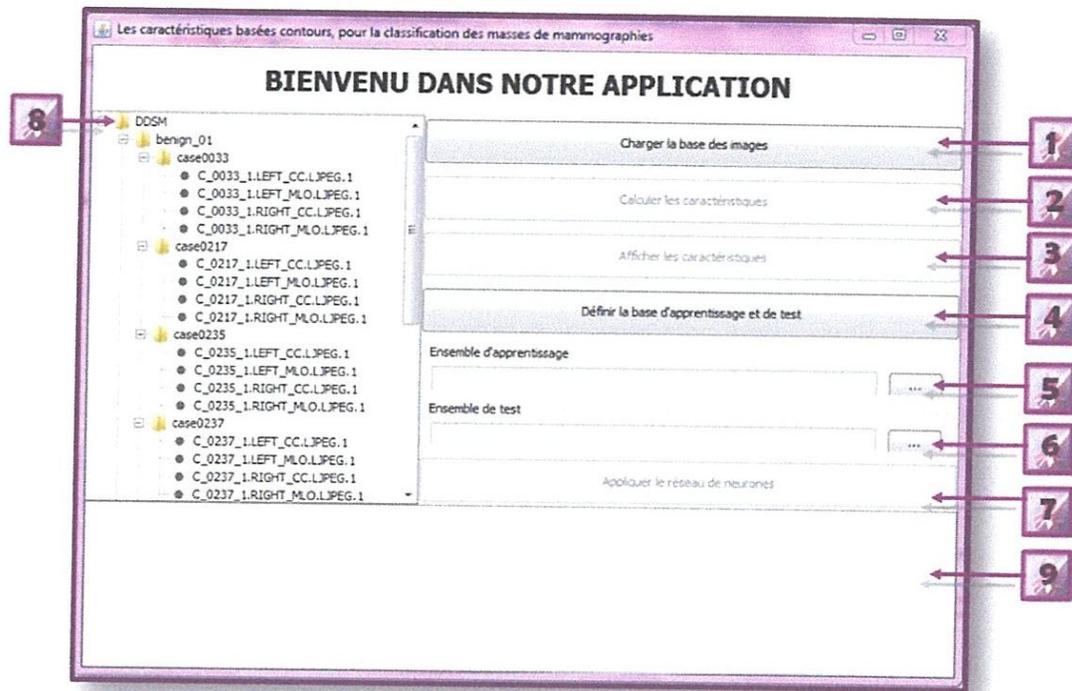


FIGURE 4.10 – interface principale de notre application.

1. Bouton pour charger la base des images :

Lorsqu'on clique sur ce bouton on charge tout les contours des masses à partir des fichiers .overlay qui se traduit en ROI (region of interest).

2. Bouton pour calculer nos caractéristiques :

Il calcule les caractéristiques des méthodes SGLDM, RDM et

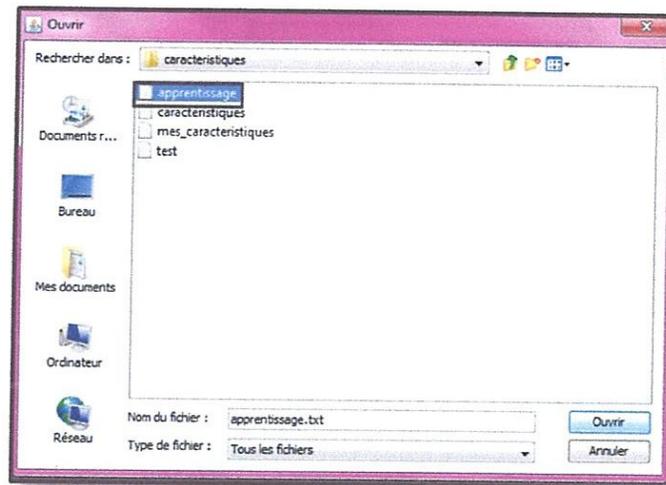


FIGURE 4.12 – Chargement des caractéristiques d'apprentissage.

6. **Bouton pour accéder au fichier qui contient les caractéristiques de la base de test :**

nous sélectionnons un fichier texte qui contient les caractéristiques de la base de test.

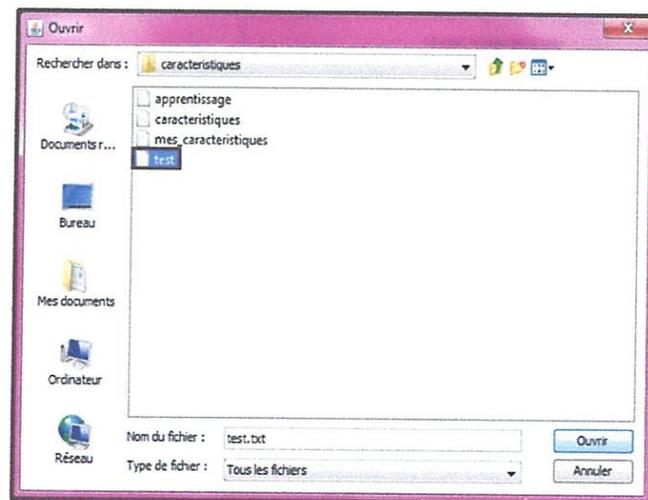


FIGURE 4.13 – Chargement des caractéristiques de teste.

l'intensité, le max de gradient, la taille de la masse et la courbure.

3. **Bouton pour un simple affichage de nos caractéristiques calculées :**

Un simple affichage des valeurs de toutes les caractéristiques de chaque masse.

4. **Bouton pour choisir le pourcentage de la base d'apprentissage et de test :**

L'image suivante montre le choix que nous avons.

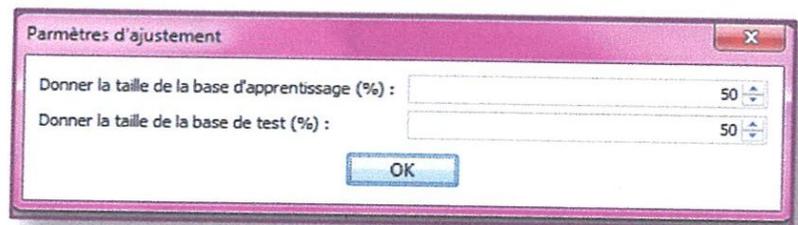


FIGURE 4.11 – Paramètres d'ajustement.

5. **Boutons pour accéder au fichier qui contient les caractéristiques de la base d'apprentissage :**

Nous sélectionnons un fichier texte qui contient les caractéristiques de la base d'apprentissage.

7. Bouton pour appliquer le réseau de neurone

8. L'espace qui contient la base d'images DDSM.

La base contient des dossiers et dans chaque dossier il y a des cas qui sont composé de 4 images.

Nous montrons dans l'image qui suit un exemple d'une image mammographique qui contient deux boutons (afficher l'anomalie et afficher le ROI) avec l'affichage de la densité, l'intensité, max de gradient, la taille et la courbure de la masse.

- Le bouton **Afficher l'anomalie** nous montre où se situe la masse exactement se situ la masse dans l'image.

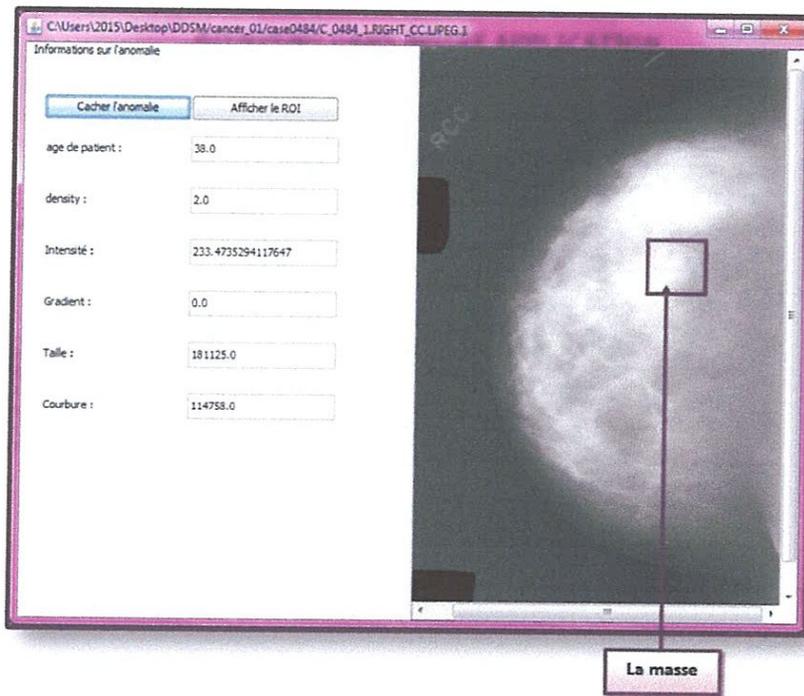


FIGURE 4.14 – Affichage d'anomalie

- Le bouton **Afficher le ROI** nous illustre la région d'intérêt.

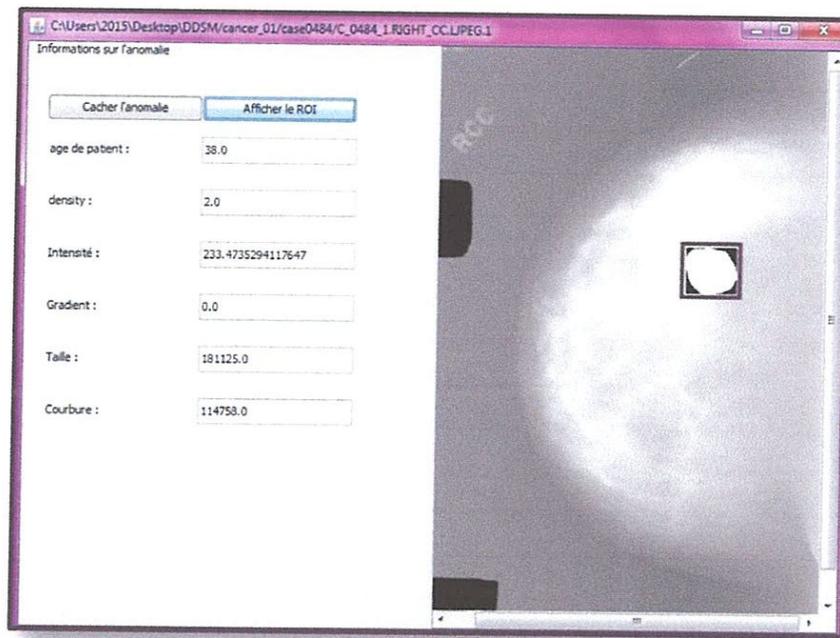


FIGURE 4.15 – Affichage de ROI

9. Une zone pour afficher les résultats de réseau de neurone :

On affiche notre résultat de classification dans cette zone avec le taux de reconnaissance.

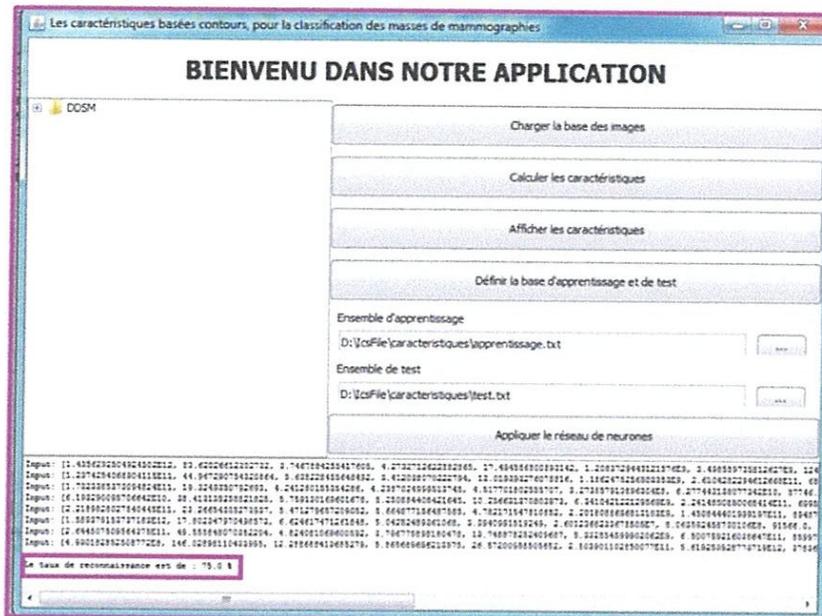


FIGURE 4.16 – Affichage de résultat

why?

4.5 Résultats et discussion

Pour notre base nous avons utilisées 25 masses entre bénigne et maligne et nous avons choisi 70% pour la base d'apprentissage ça veut dire 17 masses et 30% pour la base de test alors 8 masses.

Le taux de reconnaissance = le nombre de décisions correctes / le nombre de décisions total

- Nos résultats :

- Les cas corrects = 6
- Les cas incorrects = 2
- Taux de reconnaissance = 75%

Très faibles
juste 8!!

D'après le résultat de notre application il n'y a moins de cas incorrects qui s'approuve que notre système est convaincant.

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.

Les résultats obtenus avec cet algorithme sont intéressants.

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 qui sont la méthode SGLDM et RDM et chacune d'elle a ses propres caractéristiques que nous avons cité dans le chapitre 2.

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 trois étapes, l'objectif de la première étape est de créer une base de données relationnelle basée sur la technologie JPA (Java Persistence API) pour la gestion de la base mammographique DDSM, cette phase facilite la tâche d'extraction d'informations relatifs aux clichés mammographiques et aux anomalies, dans la deuxième phase un ensemble de caractéristiques sont calculées à travers des méthodes connues présentés dans le chapitre 2 (SGLDM et RDM) pour renforcer leur pouvoir en décrire les caractéristiques pertinentes des masses mammaires, et la troisième 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 sont satisfaisants au vu des résultats présentés dans la littérature et confirmés par des médecins radiologues. Les résultats sont encourageants de cette étude.

Perspectives

Ce mémoire a permis de dégager quelques perspectives comme l'amélioration du taux de classification (en minimisent le taux d'erreur) et la comparaison des résultats obtenus avec d'autres classifieurs.

Nous devons finir la partie décisionnelle de l'algorithme en ce qui concerne la classification des anomalies détectés selon Bi-RADS de 1 à 5.

Bibliographie

- [1] Kim, J. K, Park, J. M, Song, K. S, et Park, H. W. Adaptive mammographic image enhancement using first derivative and local statistics. *IEEE Transactions on Medical Imaging*, 16(5) :495–502,1997.
- [2] Petrou, M. and Garcia-Sevilla, P. *Image processing : dealing with texture*. Wiley,. 2006.
- [3] Haralick, R. M ,Dinstein, I, et Shanmugam, K. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6) :610–621,1973.
- [4] Aristofanes C. Silva, Anselmo C. Paiva, Paulo C.P. Carvalho, et Marcelo Gattass , "Semivariogram and SGLDM Methods Comparison for the Diagnosis of Solitary Lung Nodule," Springer, pp. 479–486,2005.
- [5] Belal Kamal Elfarra, *Mammogram Computer-Aided Diagnosis*,2012.
- [6] P. Mohanaiah, P. Sathyana rayana, L. GuruKumar, image texture feature extraction using glcm approach, *International Journal of Scientific and Research Publication*,2013.
- [7] Ruaa MohammedHamza Dr. Tawfiq A. Al-Assadi, Genetic Algorithm to find optimal GLCM features, 2002.
- [8] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E. L., et Alim, O. A. Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6) :799–810,1997.
- [9] Sahiner, B. S., Chan, H. P., Petrick, N., Helvie, et Hadjiiski, L. M. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Med. Phys.*, 28(7) :1455–1465,2001.
- [10] Zhang, D. et Lu, G. Review of shape representation and description techniques. *Pattern Recognition*. 37 :1–19,2004.
- [11] Cheikhrouhou, I., Djemal, K., Sellami, D., Derbel, N, et Maaref, H . Abnormalities description for breast cancer recognition. In *The First International Conference on E-Medical Systems E-Medisys 07*. Pages 198–205,2007.

- [12] Mingqiang, Y., Kidiyo, K, et Joseph, R. A survey of shape feature extraction techniques. *Pattern Recognition*, pages 43–90,2008.
- [13] Cheikh rouhou, I., Djemal, K., Sellami, D., Maaref, H , et Derbel, N. Empirical descriptors evaluation for mass malignity recognition. In *TheFirst International Workshop on Medical Image Analysis and Description for Diagnosis Systems - MIAD'09*. Pages 91–100,2009.
- [14] Peura, M. et Iivarinen, J. Efficiency of simple shapedescriptors. In *Third International Workshop on VisualForm*, pages 443–451,1997.
- [15] Nguyen, T. et Debled-Rennesson, I. Curvature estimation in noisy-curves. In *International conference on Computer Analysis of Images and Patterns*, volume 4673, pages 474–481,2007.
- [16] Coeurjolly, D., Miguet, S., and Tougne, L . Discretecurvature based on osculatingcircle estimation. *Proc. Int. workshop VisualForm.*, 2059 of *Lecture Notes in Computer Science*, Springer : 303–302,2001.
- [17] Kerautret, B. and Lachaud, J . Robust estimation of curvaturalong digital contours with global optimization. In *International conference on discretegeometry for computer imagery*, volume 4992 of *LNCS*, pages 334–345,2008.
- [18] Malgouyres, R., Brunet, F., and Fourey, S. Binomial convolutions and derivatives estimations from noisy discretizations. In *International conference on discretegeometry for computer imagery*. Volume 4992 of *LNCS*, pages 370–379,2008.
- [19] Zora K. Brown Harold P. Freeman et Elizabeth Platt « 100 Question –réponses le cancer du sein »copyright 2007.
- [20] Lanyi, M . Morphologi canalysis of micro calcifications. *Early Breast Cancer*, edJ.Zander and J. Baltzer, Springer-Verlag, Berlin,1985.
- [21] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., and Sickles., E. A .*American college of radiology (breast imaging reporting and data system)*. Troisième édition française réalisée par SFR (Société Française de Radiologie),2003.
- [22] Wolfe, J. N. Breast patterns as an index ofrisk for developping breast cancer. *American Journal of Roentgenology* , 126(6) :1130–1137,1976.
- [23] Le Gal, M., Chavanne, G., and Pellier, D. Valeur diagnostique des micro calcifications groupées découvertes par mammographie. À propos de 227 cas. *Bull Cancer*, 71 :57–64,1984.
- [24] Feig, S. A., D'Orsi, C. J., Hendrick, R. E., Jackson, V. P., Kopans, D. B., Monsees, B., Sickles, E. A., Stelling, C. B., Zininger, M.,

- and Wilcox-Buchalla, P. American college of radiology guide lines for breast cancer screening. American Journal of Roentgenology, 171(1) :29–33,1998.
- [25] Clément Chatelain, reconnaissance des formes : application aux documents, volume 11, 2015.
- [26] B. Dubuisson, 'Diagnostic et Reconnaissance des Formes'. Traité des nouvelles technologies. Série diagnostic et Maintenance. Hermès, Paris, 1990 .
- [27] Laurent HEUTTE. Reconnaissance de Formes Processus de RdF, volume 9.
- [28] Richard lepage. Reconnaissance d'algues toxiques par vision artificielle et réseau de neurones. volume 202, 2004.
- [29] Muhammad ARIF, usion de Données : Ultime Étape de Reconnaissance de Formes, Applications à l'identification et à l'authentification. volume 272, 2005.
- [30] Nicholas Journet, Introduction au traitement d'images : Reconnaissance des formes, volume 26, 2011.
- [31] James L. Crowley. Reconnaissance de Forme Statistique. volume 20, 2001.
- [32] Swarup Medasani Jaeseok Kimu, Raghu Krishnapuram. An overview of membership function generation techniques for pattern recognition. International Journal of Approximate Reasoning 19, volume 27, 1998.
- [33] Seridi Hamid , reconnaissance de formes , volume 15 , 2009.
- [34] Rolf Ingold, reconnaissance des formes introduction , University of Fribourg, volume 29, 2006.
- [35] Yousri Kessentini. Reconnaissance De Formes Classification « statistique » des formes, volume 90, 2014.
- [36] Manel Zribi, Younes Boujelbene, Les réseaux de neurones un outil de sélection de variables : Le cas des facteurs de risque de la maladie du cancer du sein, volume 10, 2012.
- [37] Bernard Gosselin. Application de réseaux de neurones artificiels a la reconnaissance automatique de caractères manuscrits. Dissertation originale présentée pour l'obtention du grade de Docteur en Sciences Appliquées, volume 231, 1993.
- [38] Yvonne Delevoye-Turrell. Introduction aux réseaux de neurones artificiels, volume 48, 2015.

- [39] Ait Skourt Brahim, Keheli Adnane, Namir Yassine, Ragui Achraf, Sadouk Amine. LE perceptron multicouche , volume 7, 2015.
- [40] Marc Parizeau. Le perceptron multicouche et son algorithme de rétro-propagation des erreurs, volume 8, 2004.
- [41] A. Belaïd. Reconnaissance des formes, volume 43, 1992.
- [42] Richard lepage , RECONNAISSANCE D'ALGUES TOXIQUES PAR VISION ARTIFICIELLE ET RÉSEAU DE NEURONE, Memoire de recherche, 2004.
- [43] Abid L, « cancer epidemiology in algeria : best use of cancer registers » ; african journal of cancer, Vol 1, Springer, pp 98-103, 2009.
- [44] University of South Florida. Digital Database for Screening Mammography (DDSM), 2008.
- [45] American College of Radiology, "Breast imaging reporting and data system : BI-RADS atlas," American College of Radiology, Reston, atlas 2003.

Webographie

[A]

http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_sein_pm

[B]

https://java.com/fr/download/faq/whatis_java.xml

[C]

<http://righthere-rightnow.over-blog.com/article-microcalcifications-cancer-du-sein-85301068.html>

[D]

http://www.aly-abbara.com/echographie/biometrie/scores/microcalcifications_classifications.html