



UNIVERSITÉ SULTAN MOULAY SLIMANE  
Faculté des Sciences et Techniques  
Béni Mellal



*Centre d'Etudes Doctorales : Sciences et Techniques*

*Formation doctorale : Mathématiques et Physique Appliquées*

## THÈSE

Présentée par

**Khaddouj TAIFI**

Pour l'obtention du grade de

**Docteur**

*Spécialité : Physique-Informatique*

---

### Détection et classification des microcalcifications mammaires pour le diagnostic du cancer du sein

---

Soutenue le 25 février 2017  
Devant la commission d'examen :

Pr. Abdelouhab ZEROUAL	: Professeur à la FSSM, Marrakech	Président
Pr. Moha M'rabet HASSANI	: Professeur à la FSSM, Marrakech	Rapporteur
Pr. Najlae IDRISSE	: Professeur à la FST, Béni Mellal	Rapporteur
Pr. Benaceur OUTTAJ	: Professeur à la FSJES, Rabat	Rapporteur
Pr. Belaid BOUIKHALNE	: Professeur à la FP, Béni Mellal	Examineur
Pr. Mohamed FAKIR	: Professeur à la FST, Béni Mellal	Co-encadrent
Pr. Said SAFI	: Professeur à la FP, Béni Mellal	Encadrent

# Liste des publications

## Chapitre

1. K. TAIFI, S. SAFI, M.FAKIR, R.AHDID, A Comparison of Wavelet and Steerable Pyramid for Classification of Microcalcification in Digital Mammogram, Springer International Publishing Switzerland 2016, Proceedings of the Mediterranean Conference on Information and Communication Technologies 2015, Lecture Notes in Electrical Engineering 380, DOI 10.1007/978 – 3 – 319 – 30301 – 7\_45

## Journaux internationaux :

2. K. TAIFI, S. SAFI, M.FAKIR, A. ELBALAOUI, Breast cancer diagnosis system based on wavelet analysis and nearest neighborhood classifier, International Journal of Computer Vision and Image Processing (IJCVIP),page 1-16,vol 4,2014.
3. K. TAIFI, R. AHDID, M.FAKIR, S. SAFI, A Hybrid the Nonsubsampled Contourlet Transform and Homomorphic Filtering for Enhancing Mammograms, TELEKOMNIKA Indonesian Journal of Electrical Engineering, Vol. 16, No. 3, December 2015, pp. 539-545.

## Conférences internationales :

7. K. TAIFI, S. SAFI, R. AHDID, M.FAKIR, A Comparison of wavelet and Steerable Pyramid for Classification of Microcalcification in digital mammogram, Mediterranean Conference on Information and Communication Technologies 2015 May 7-9, 2015 Saïdia, Morocco MedICT'15
8. K. TAIFI, R.AHDID, S. SAFI, Comparing the Performance of Image Segmentation Methods to Detect Mass in Digital Mammography, International Conference on Multimedia Computing and Systems(ICMCS'14) IEEE May 14-16, 2014,Marrakech, Morocco,
9. K. TAIFI, S. SAFI, M.FAKIR, Detection of Microcalcifications and classification, 8th International Conference on Intelligent Systems : Theories and Applications (SITA' 13) IEEE, May 08-09 2013, Rabat, Morocco
10. K. TAIFI, S. SAFI, M.FAKIR, A. ELBALAOUI, Breast cancer diagnosis system based on wavelet analysis and nearest neighborhood classifier, 3th International Sym-

- posium on Amazigh Language Processing (SITACAM'2013), May 2-4, 2013, Béni Mellal
11. K. TAIFI, S. SAFI ,B. BOUIKHALNE, A. BOUMEZOUG, M. FAKIR, Microcalcifications Detection using the Homomorphic Filtering and Wavelet, The 3rd International Conference on Multimedia Computing and Systems (ICMCS'12) IEEE, May 10-12, 2012, Tanger, Morocco.
  12. K. TAIFI, S. SAFI, M.FAKIR, R. AHDID, A Hybrid the Nonsampled Contourlet Transform and Homomorphic Filtering for Enhancing Mammograms, Second International Conference on Business Intelligence(CBI'15), Beni Mellal, Morocco, April 23-25, 2015
  13. K. TAIFI, S. SAFI, M.FAKIR, R. AHDID, Enhancement Mammograms with Non-sampledContourelet Transformation and Feature Extraction for classification Second International Conference on Business Intelligence(CBI'15), Beni Mellal, Morocco, April 23-25, 2015
  14. K. TAIFI, S. SAFI, M. FAKIR, Mammogram classification using statistical feature and gray-level co-occurrence matrix for detection of breast , 2ème Edition des Journées Doctoriales(JDoc'15) 2015
  15. K. TAIFI, S. SAFI, M.FAKIR, A. ELBALAOUI, R.AHDID, Suppression Pectoral Muscle and Artifacts in digital Mammograms, International Conference on Business Intelligence (CBI' 14), Beni Mellal, Morocco, April 29-30, 2014.

# Dédicaces

À la mémoire de mon père, qu'Allah ait pitié de son âme  
À celle que j'aime plus que tout au monde, ma très chère mère  
À mes chères frères et soeurs  
À la famille TAIFI  
À auxquels, je dois ce que je suis

<<La recherche procède par des moments distincts et durables, intuition, aveuglement, exaltation et fièvre. Elle aboutit un jour à cette joie, et connaît cette joie celui qui a vécu des moments singuliers.>>  
*Albert Einstein*

# Remerciements

Les travaux présentés dans cette thèse ont été réalisés au sein de l'unité de Formation de Recherche : Laboratoire de Traitement de l'Information et Aide à la Décision (TIAD) du département Informatique de la Faculté des Sciences et techniques de Béni Mellal.

Je veux spécialement remercier mon directeur de recherche Monsieur S.SAFI, Professeur à la Faculté Polydisciplinaire de Béni Mellal et lui exprimer ma gratitude pour sa disponibilité et son soutien pendant ces années de recherche.

Je tiens à exprimer toute ma profonde gratitude à mon co-encadrent Monsieur M. FAKIR, Professeur à la Faculté des Sciences et techniques de Béni Mellal, pour son aide et pour sa sympathie, son ouverture d'esprit et son sens critique.

Mes remerciements s'adressent également à l'honorable jury d'avoir accepté d'examiner mon mémoire de thèse. Merci à Monsieur Abdelouhab ZEROUAL, Professeur à la FSSM de Marrakech, d'avoir accepté de présider le jury de cette thèse et à Messieurs Moha M'rabet HASSANI, Professeur à la FSSM de Marrakech, Benaceur OUTAJ, Professeur à la FSJES Rabat et Madame Najlae IDRISSE, Professeur à la FST de Béni Mellal, d'avoir accepté de rapporter ce manuscrit. Je remercie aussi Monsieur Belaid BOUIKHALNE, Professeur à la FP de Béni Mellal, d'avoir accepté d'examiner ce travail.

Je tiens à remercier mes parents, mes soeurs et mes frères pour leur soutien durant toutes ces années d'études.

Je voudrais remercier Monsieur Ahmed BOUMEZOUG, Meriem EL HARFI, mes amis, les collègues du laboratoire TIAD et tous ceux qui m'ont aidé de près ou de loin pour élaborer cette thèse.

# Résumé

Le cancer du sein est la principale cause de mortalité qui touche les femmes à travers le monde. Le manque de dépistage ou diagnostic tardif peut s'avérer inutile à des stades avancés de la maladie. Pour cela, un diagnostic précoce permettra d'anticiper la présence et/ou l'évolution du cancer du sein. Pour ce faire, les radiologues ou médecins spécialistes ont recours à la mammographie qui consiste en un ensemble d'examen radiologiques des tissus intérieurs du sein basés sur les rayons X. Cette technique d'imagerie médicale met en évidence les anomalies qui puissent exister telles que les microcalcifications et les masses.

Néanmoins, malgré les avantages fournis par la mammographie, la difficulté d'interprétation reste la limite majeure de cette technique.

Dans cette thèse, nous proposons un système d'analyse d'images mammographiques intégrant les techniques de traitement d'images pour la détection et l'interprétation des lésions liées au cancer du sein. Le but de ce système est la prise de décision optimale des images mammographiques étudiées afin de classifier les tissus normaux des tissus anormaux.

Dans un premier temps, nous proposons d'extraire le muscle pectoral par la méthode de ligne de partage d'eau (LPE) morphologique. Cette méthode est précédée d'une étape de prétraitement pour extraire la zone du sein de l'image mammographique, en supprimant toutes les types des étiquettes existantes dans l'image. De cette manière, la segmentation du muscle pectoral peut se faire d'une manière fiable et simple.

L'algorithme utilisé pour extraire les microcalcifications procède dans un premier temps par un prétraitement basé sur le filtrage homomorphique et des algorithmes de décomposition multi-échelle comme : le filtre de gabor, les ondelettes et les Contourlets non sous échantillonnée

La phase suivante concerne l'extraction des caractéristiques pour la classification des tissus normaux et anormaux dans la mammographie en se basant sur les réseaux de neurones artificiels, les machines à vecteurs de support et l'algorithme de k-plus proche voisin.

Les résultats expérimentaux obtenus à partir de ce système se voient bénéfique pour la détection automatique du cancer du sein et à l'aide à une prise de décision plus fiable.

**Mots-clés :** Cancer du sein, Diagnostic Assisté par Ordinateur ( $DAO_x$ ), Mammographie, Segmentation, Classification.

# Abstract

Breast cancer is the leading cause of mortality among women worldwide. Lack of screening or late diagnosis may be unnecessary at later stages of the disease. For this, an early diagnosis will make it possible to anticipate the presence and/or the evolution of the breast cancer. To do this, radiologists or medical specialists use mammography which consists of a set of radiologic examinations of the internal tissues of the breast based on X-rays. This medical imaging technique highlights the anomalies that may exist such as microcalcifications and masses.

Nevertheless, despite the advantages provided by mammography, the difficulty of interpretation remains the major limit of this technique.

In this thesis, we propose a mammography image analysis system incorporating technical image processing for the detection and interpretation of lesions associated with these diseases. The purpose of this system is the optimal classification between normal and abnormal tissues in a mammogram.

First, we propose to extract the pectoral muscle is based on a morphological watershed. The presented method is preceded by a pretreatment step to remove the breast area of the mammographic image, removing all types of mammography film labels. In this way, the segmentation of the pectoral muscle can be easily removed.

The algorithm used to extract abnormal microcalcification contains a method of pretreatment based on homomorphic filtering and multi-scale decomposition algorithms like : the gabor filter, wavelets and nonsubsampling contourlet transform (NSCT).

The next phase containing feature extraction for classification of normal and abnormal tissue in mammography is based on an Artificial Neural Network (ANN), K Nearest Neighbor (KNN) and Support Vector Machine (SVM).

The experimental results obtained from this system are beneficial to the automatic detection of breast cancer and for more reliable decision.

**Keywords :** Breast cancer, Computer Aided Diagnosis systems ( $CAD_x$ ), Mammogram, Segmentation, Classification.

# Table des matières

Liste des publications	ii
Dédicaces	iv
Remerciements	v
Résumé	vi
Abstract	vii
Table des figures	xi
Liste des tableaux	xiv
Abréviations	xv
<b>Introduction</b>	<b>1</b>
1. Objectifs et contributions	1
2. Structure de la thèse	2
<b>1 Cancer du sein</b>	<b>4</b>
Introduction	4
1. Anatomie du sein	5
2. Dépistage	6
3. Modalités d'imagerie	7
3.1. L'échographie	7
3.2. La mammographie	8
3.3. L'imagerie par résonance magnétique	9
4. Les tumeurs du sein	10
4.1. Les tumeurs mammaires non cancéreuses	10
4.2. Les tumeurs mammaires cancéreuses	11
5. Les pathologies mammaires	11
5.1. Les masses	11
5.2. Les microcalcifications	12
5.2.1. L'étude morphologique des microcalcifications	12
5.2.2. Distribution	17
6. Symptômes	18



7.	Facteurs de risque . . . . .	19
8.	Base de données mammographies . . . . .	19
8.1.	Base de données MIAS . . . . .	20
8.2.	Base de données DDSM . . . . .	20
	Conclusion . . . . .	21
<b>2</b>	<b>Technique de segmentation : application aux images mammographiques</b>	<b>22</b>
	Introduction . . . . .	22
1.	Méthodes de prétraitement . . . . .	22
1.1.	Les différents types de filtres . . . . .	23
1.1.1.	Filtre passe-bas . . . . .	23
1.1.2.	Filtre passe-haut . . . . .	24
1.1.3.	Filtre médian . . . . .	25
2.	Méthodes de segmentation . . . . .	26
2.1.	Segmentation par seuillage . . . . .	26
2.2.	Segmentation par extraction de régions . . . . .	26
2.2.1.	Segmentation par croissance de régions . . . . .	27
2.2.2.	Segmentation par division et fusion . . . . .	27
2.3.	Segmentation par détection de contours . . . . .	27
3.	Méthodes de détection des microcalcifications . . . . .	28
3.1.	Filtre de Gabor . . . . .	30
3.2.	Ondelettes . . . . .	30
3.2.1.	La transformée en ondelettes continue . . . . .	30
3.2.2.	La transformée en ondelettes discrète . . . . .	31
3.2.3.	Les familles de la transformée en ondelettes . . . . .	36
3.3.	La transformée en Contourlet . . . . .	37
3.4.	La transformée en contourlet non sous échantillonnée . . . . .	37
3.4.1.	Pyramide non sous échantillonnée . . . . .	38
3.4.2.	Bancs de filtres directionnels non-sous-échantillonnés . . . . .	39
3.5.	La morphologie mathématique . . . . .	42
4.	Segmetation du muscle pectoral . . . . .	43
	Conclusion . . . . .	43
<b>3</b>	<b>Classifications des images mammographiques</b>	<b>45</b>
	Introduction . . . . .	45
1.	Les descripteurs de la couleur . . . . .	46
2.	Les descripteurs de la forme . . . . .	46
3.	Les descripteurs de la texture . . . . .	46
3.1.	Analyse des textures par les méthodes statistiques . . . . .	46
3.1.1.	Technique de Tamura . . . . .	46
3.1.2.	Matrices de co-occurrences . . . . .	48
3.2.	Analyse des textures par les méthodes fréquentielles . . . . .	50
3.2.1.	L'énergie des ondelettes . . . . .	50
3.2.2.	Le descripteur GIST . . . . .	50
4.	L'apprentissage et la classification . . . . .	52
4.1.	Les types de la classification . . . . .	52

4.1.1.	Machine à Vecteurs de Support (SVM) . . . . .	54
4.1.2.	K Plus Proches Voisins (KPPV) . . . . .	57
4.1.3.	Réseaux de Neurons (RN) . . . . .	58
5.	Evaluation de la performance de classification . . . . .	63
	Conclusion . . . . .	65
<b>4</b>	<b>Contributions : système de segmentation et classification des images mam-</b>	<b>66</b>
	<b>mographiques</b>	<b>66</b>
	Introduction . . . . .	66
1.	Segmentation du muscle pectoral . . . . .	67
1.1.	Prétraitement de l'image . . . . .	68
1.1.1.	Suppression du bruit . . . . .	69
1.1.2.	Suppression des artefacts Radiopaques . . . . .	69
1.2.	Suppression des muscles pectoraux . . . . .	70
1.3.	Evaluation des performances : . . . . .	75
1.4.	Résultats et discussions . . . . .	76
2.	Détection des microcalcifications . . . . .	79
2.1.	Prétraitement . . . . .	79
2.1.1.	Extraction des régions d'intérêt (ROI) . . . . .	79
2.1.2.	Rehaussement du contraste . . . . .	81
2.1.3.	Filtrage par un filtre haute fréquence . . . . .	81
2.2.	Détection des microcalcifications . . . . .	82
2.2.1.	Analyse par multirésolution des transformée en ondelettes . . . . .	82
3.	Classification des images mammographiques . . . . .	85
3.1.	Classification des images normales et anormales avec les ondelettes . . . . .	85
3.1.1.	Effet du prétraitement sur les résultats de la décision . . . . .	85
3.1.2.	Effet des niveaux de résolution sur les résultats de la décision . . . . .	86
3.1.3.	Effet des classifieurs sur les résultats de la décision . . . . .	90
3.2.	Classification des images normales et anormales avec les contourlets non sous échantillonnées . . . . .	90
3.2.1.	Approche hybride combinant la transformation de Contour- let non sous échantillonné et le filtre homomorphique . . . . .	90
3.2.2.	Techniques d'évaluation d'amélioration du contraste pour mammographique . . . . .	91
3.2.3.	Résultats expérimentaux . . . . .	92
3.3.	La phase de classification . . . . .	93
	Conclusion . . . . .	94
	<b>Conclusion et perspectives</b>	<b>95</b>
	<b>Bibliographie</b>	<b>97</b>

# Table des figures

1	Schéma du processus du traitement automatique des images mammographiques	2
1.1	Les 4 cancers les plus fréquents chez la femme.	4
1.2	Anatomie du sein [1]	6
1.3	Echographie présentant :a) des microcalcifications, b) des microkystes	8
1.4	Image mammographique d'un sein.	8
1.5	IRM d'un sein	9
1.6	Formes des masses.	11
1.7	Microcalcifications cutanées	12
1.8	Microcalcifications rondes	13
1.9	Microcalcifications à centres clairs	13
1.10	Microcalcifications dystrophiques	13
1.11	Microcalcifications amorphes	14
1.12	Microcalcifications vasculaires	14
1.13	Microcalcifications en bâtonnets	15
1.14	Microcalcifications d'oeuf	15
1.15	Microcalcifications de suture	15
1.16	Microcalcifications polymorphes	16
1.17	Microcalcifications grossières	16
1.18	Microcalcifications à type de lait calcique	17
1.19	Microcalcifications linéaires	17
1.20	Les différentes distributions des microcalcifications : a) Mcs diffuses ou éparées : b) Régionale, c) Cluster, d) Linéaire et e) Segmentaire.	18
2.1	Procédure de diagnostic assistée par ordinateur	22
2.2	Le principe du filtre Médian	25
2.3	Principe de la procédure d'immersion : l'immersion de ses minima régionaux, tout en empêchant la fusion des eaux provenant de différentes sources, partitionne l'image en les bassins versants (catchment bassins) et les lignes de partage des eaux (watersheds)[2]	28
2.4	Décomposition du signal $x(n)$	32
2.5	Schéma bloc de la décomposition par ondelettes en trois itérations	34
2.6	Processus d'analyse multi-résolution à deux dimensions pour une itération	34
2.7	Décomposition d'une image en trois itérations	35
2.8	Processus de reconstruction à deux dimensions pour une itération	36
2.9	Quelques familles d'ondelettes	37

2.10	Transformation en contourlets non sous-échantillonnée : a) Schéma bloc de l'implantation de la NSCT à l'aide d'un banc de filtre pyramidal directionnel non sous échantillonné, b) Banc de filtre directionnel [3]. . . . .	38
2.11	Pyramide de décomposition non sous échantillonnée à 3 niveaux. . . . .	39
2.12	Bancs de filtres directionnels non-sous-échantillonnés construits avec des bancs de filtres en éventail à deux-canaux. . . . .	40
2.13	Filtres sur-échantillonnés par une matrice quinconce Q . . . . .	41
2.14	Le bloc de diagramme de la NSCT . . . . .	41
3.1	Schéma de principe pour le calcul et l'extraction du descripteur de GIST . . . . .	51
3.2	Extrait de la classification taxinomique de Linné. . . . .	53
3.3	L'apprentissage supervisé. . . . .	53
3.4	Les modèles des SVM : a) cas linéairement séparable, b) non linéairement séparable. . . . .	54
3.5	L'hyperplan H optimal, vecteurs supports et marge maximale. . . . .	55
3.6	Transformation des données dans un espace de grande dimension . . . . .	57
3.7	Principe de classification par KPPV . . . . .	58
3.8	Représentation de la forme biologique d'un neurone . . . . .	58
3.9	Le modèle de neurone formel . . . . .	59
3.10	Les fonction de transfert pour le neurone artificiel. . . . .	60
3.11	Exemple d'un perceptron monocouche . . . . .	61
3.12	Perceptron multicouches avec une seule couche cachée . . . . .	62
4.1	Système de travail . . . . .	66
4.2	Images mammographiques : a) image (mdb012), b) son histogramme, c) image (mdb209), d) son histogramme . . . . .	67
4.3	Schéma proposé de segmentation du muscle pectoral . . . . .	68
4.4	Bruits de numérisation en images mammographiques . . . . .	69
4.5	Images de mammographie après suppression du bruit par le filtre Gaussien . . . . .	70
4.6	Image (mdb012) mammographique initiale . . . . .	71
4.7	Images résultantes avec : a) seuillage $T = 0.13$ , b) puis opérations morphologiques $R=55$ . . . . .	71
4.8	Image mammographique après prétraitement . . . . .	71
4.9	Principales étapes de segmentation du muscle pectoral : a) Image originale, b) prétraitement, c) gradient . . . . .	72
4.10	Principales étapes de segmentation du muscle pectoral : a) LPE, b) morphologie mathématique, c) maxima régionaux . . . . .	72
4.11	Principales étapes de segmentation du muscle pectoral : a) LPE de gradient modifié, b) détermination du muscle pectoral par LPE, c) segmentation du muscle . . . . .	73
4.12	Segmentation du muscle pectoral(a), b) binarisation du muscle pectoral, c) suppression du muscle pectoral . . . . .	73
4.13	Résultats de la détection du muscle pectoral par la méthode Jayasree Chakraborty [4] . . . . .	74
4.14	Segmentation du muscle pectoral par la méthode proposée . . . . .	74

---

4.15	Segmentation du muscle pectoral de l'image mdb199 : a) manuellement, b) par la méthode Jayasree Chakraborty, c) par la méthode de Ferrari, d) par la méthode de Kwok, et e) par la méthode proposée . . . . .	75
4.16	Résultat de l'évaluation avec vrai positif et de faux positif . . . . .	77
4.17	Résultat de l'évaluation avec sensibilité et spécificité . . . . .	77
4.18	Résultat de l'évaluation avec l'exactitude et la précision . . . . .	78
4.19	Résultat de l'évaluation avec le taux d'équilibre de classification et le taux d'erreur . . . . .	78
4.20	Image mammographique . . . . .	79
4.21	Extraction d'échantillons en utilisant différentes "carrés" tailles de ROI . . . . .	80
4.22	Extraction du ROI 256 * 256 de deux images différentes . . . . .	80
4.23	Résultat de normalisation : a) Image originale, b) Normalisation, c) Histogramme de l'image originale, d) Histogramme de l'image normalisée . . . . .	81
4.24	Image après filtre homomorphique . . . . .	82
4.25	Les approximations et les détails . . . . .	83
4.26	Annulation des détails . . . . .	83
4.27	Annulation des approximations . . . . .	83
4.28	Segmentation des microcalcifications : a) Image originale, b) image après le seuillage, c) contour coloré de l'image (b) superposé sur l'image originale (a) . . . . .	84
4.29	Courbes de classification avec et sans prétraitement . . . . .	86
4.30	Les résultats des expériences de l'application des différentes familles d'ondelettes sur les images mammographiques. . . . .	86
4.31	Le taux de classification pour gabor . . . . .	87
4.32	Le taux de classification pour gist . . . . .	88
4.33	Le taux de classification pour tamura . . . . .	88
4.34	Le taux de classification pour les trois descripteurs . . . . .	89
4.35	Comparaison de différentes techniques d'amélioration : a) ROI de l'image originale, b) le NSCT, c) le filtrage homomorphique, d) le masquage flou, e) notre approche proposée . . . . .	91
4.36	Traçage de DSM . . . . .	92
4.37	Traçage de TBCS . . . . .	93
4.38	Traçage de TBCE . . . . .	93
4.39	Taux de classification de la méthode NSCT et la méthode proposée. . . . .	94

# Liste des tableaux

1.1 Répartition du nombre de cas des principaux types de cancer enregistrés à l'Institut National d'Oncologie (INO) de Rabat (2002 à 2007) [5] . . . . .	5
3.1 Signification de VP, VN, FP et FN de la matrice de confusion . . . . .	63
3.2 Matrice de confusion . . . . .	63
4.1 Moyenne de FP, FN, et le (Tot) de la base mini-MIAS. . . . .	76
4.2 Matrice de confusion . . . . .	85
4.3 Base des images mammographiques . . . . .	85
4.4 Evaluation des résultats avec différentes bases des images . . . . .	89
4.5 Evaluation des résultats avec différents classifieurs . . . . .	90

# Abréviations

<b>Mcs</b>	Microcalcifications
<b>MIAS</b>	Mammography <b>I</b> mage <b>A</b> nalysis <b>S</b> ociety
<b>DDSM</b>	<b>D</b> igital <b>D</b> atabase for <b>S</b> creening <b>M</b> ammography
<b>INO</b>	<b>I</b> nstitut <b>N</b> ational d' <b>O</b> ncologie
<b>IRM</b>	Imagerie par <b>R</b> ésonance <b>M</b> agnétique
<b>LPE</b>	<b>L</b> igne de <b>P</b> artage d' <b>E</b> au
<b>NSCT</b>	<b>N</b> on <b>S</b> ubsampled <b>C</b> ontourlet <b>T</b> ransform
<b>DWT</b>	<b>D</b> iscrete <b>W</b> avelet <b>T</b> ransform
<b>DAO</b>	<b>D</b> iagnostic <b>A</b> ssisté par <b>O</b> rdinateur
<b>SVM</b>	<b>M</b> achines à <b>V</b> ecteurs de <b>S</b> upport
<b>KPPV</b>	<b>K</b> - <b>P</b> lus <b>P</b> roche <b>V</b> oisin
<b>RN</b>	<b>R</b> éseaux de <b>N</b> eurones
<b>FN</b>	<b>F</b> aux <b>N</b> égatifs
<b>VN</b>	<b>V</b> rais <b>N</b> égatifs
<b>FP</b>	<b>F</b> aux <b>P</b> ositifs
<b>VP</b>	<b>V</b> rais <b>P</b> ositifs
<b>TIAD</b>	<b>T</b> raitement de l' <b>I</b> nformation et <b>A</b> ide à la <b>D</b> écision ;

# Introduction

Le cancer du sein est la principale cause de mortalité qui touche les femmes à travers le monde. Le manque de dépistage ou diagnostic tardif peut s'avérer inutile à des stades avancés de la maladie. Pour cela, un diagnostic précoce permettra d'anticiper la présence et/ou l'évolution du cancer du sein. Pour ce faire, les radiologues ou médecins spécialistes ont recours un ensemble d'examen radiologiques des tissus intérieurs du sein basés sur les rayons X. Cette technique d'imagerie médicale met en évidence les anomalies qui puissent exister telles que les microcalcifications (Mcs) et les masses.

Néanmoins, malgré les avantages fournis par la mammographie, la difficulté d'interprétation reste la limite majeure de cette technique.

Dans ce contexte, ce mémoire de recherche essaye d'apporter et d'aider à l'interprétation des images mammographiques en exploitant quelques techniques de traitement et d'analyse des images. L'objectif ainsi souhaité est de fournir, à l'aide d'un système de diagnostic assisté par ordinateur (voir figure 1) une interprétation fiable et dans des délais satisfaisants.

## 1. Objectifs et contributions

Le cancer du sein peut se présenter sur plusieurs formes ou types à savoir : les masses et les microcalcifications. Dans cette thèse, nous nous focalisons sur l'étude et l'analyse des microcalcifications. Nous proposons ainsi de suivre les étapes suivantes pour une détection précoce des premiers signes du cancer du sein et qui représentent nos principales contributions dans cette thèse :

1. Suppression des muscles et des artefacts.
2. Amélioration de la qualité des images mammographiques.
3. Détection des microcalcifications.
4. Extraction des attributs.
5. Classification.



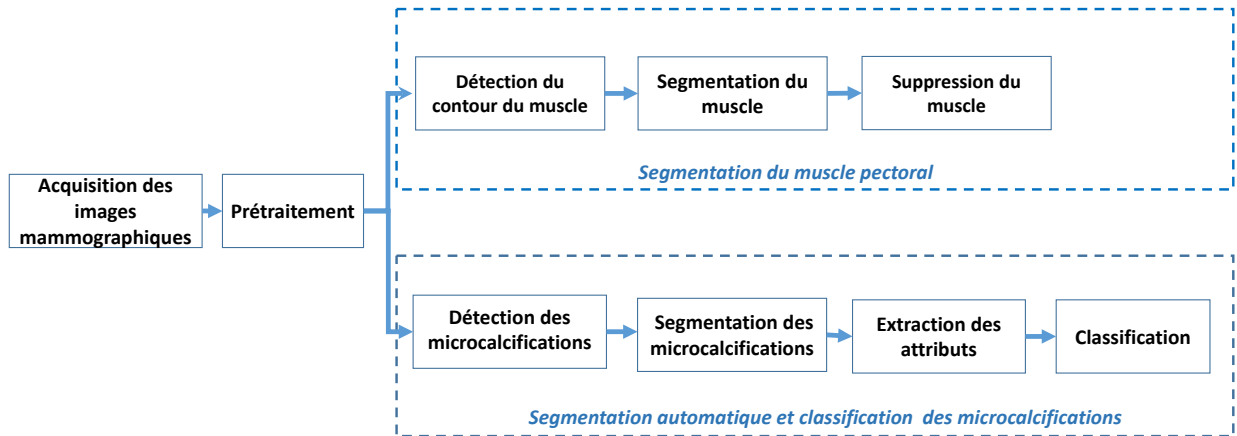


FIGURE 1 – Schéma du processus du traitement automatique des images mammographiques

## 2. Structure de la thèse

Ce mémoire est divisé en quatre chapitres.

Introduction : elle présente la problématique traitée dans cette thèse ainsi que les solutions proposées.

— Le chapitre. 1 : Cancer du sein : aspect médical

Ce chapitre présente le cancer du sein afin de justifier l'importance accordée à ce type de cancer. Les outils d'imagerie médicale permettant le dépistage et le diagnostic du cancer du sein sont abordés dans l'objectif d'identifier les avantages et les inconvénients de chaque outil. Une orientation vers la mammographie nécessite l'étude de l'anatomie du sein et son principe de fonctionnement. Ensuite, nous détaillons les caractéristiques (morphologie, distribution, forme, contour, ...) des lésions mammaires qui sont les microcalcifications et les masses à savoir, les causes et les symptômes.

— Le chapitre. 2 : Techniques de segmentation : application aux images mammographiques.

Dans ce chapitre, nous présentons un système d'analyse d'images mammographiques permettant d'une part, la segmentation des éléments principaux du sein, et d'autre part la détection des lésions mammaires. Les différentes méthodes proposées pour les segmenter sont détaillées.

— Le chapitre. 3 : Classification des images mammographiques

Ce chapitre présente d'une part, l'extraction des attributs de texture des images mammographiques, et d'autre part la classification. Les différentes méthodes proposées pour l'extraction et la classification sont détaillées.

- Le chapitre. 4 : Contributions : architecture du système de segmentation et classification des images mammographiques

Ce chapitre est le coeur de notre travail. Il est divisé en trois sections. Dans la première section, nous présentons l'algorithme de segmentation du muscle pectoral. Dans la deuxième section, nous décrivons le rehaussement des images mammographiques liées à la base des images( Mammographic Image Analysis Society (MIAS)) et les différentes méthodes proposées pour la détection des lésions lumineuses. Dans la troisième section, nous décrivons l'extraction des attributs et leurs classification. Les résultats obtenus et l'évaluation des performances des algorithmes sont également présentés.

Ce travail s'achève par une conclusion générale qui récapitule succinctement les différents travaux entrepris dans ce mémoire. Elle met en valeur les principaux résultats obtenus et donne quelques perspectives.

# Chapitre 1

## Cancer du sein

### Introduction

Actuellement, le cancer du sein est le premier cancer de la femme dans le monde et son incidence augmente. Par conséquent, la recherche et le développement d'un système d'analyse des images du sein pour l'aide au diagnostic attire l'attention de beaucoup de chercheurs.

Dans ce chapitre, nous commençons par présenter l'anatomie du sein et ses principales composantes. Ensuite, les différents types d'imagerie médicales nécessaires au diagnostic. Et, nous terminons par une synthèse des tumeurs du sein les plus connues et répandues.

Au Maroc, d'après le registre des cancers de la région du Casablanca pour l'année 2004, le cancer du sein et le cancer du col utérin constituent environ la moitié des cancers féminins et 27% de tous les cancer, tous les âges confondus [6, 7, 8, 9, 10]. Le cancer du sein représente le premier cancer de la femme, son incidence standardisée est de 30 pour 100.000 femmes par année. Il représente 36,1 % de tous les cancers féminins (voir figure 1.1).

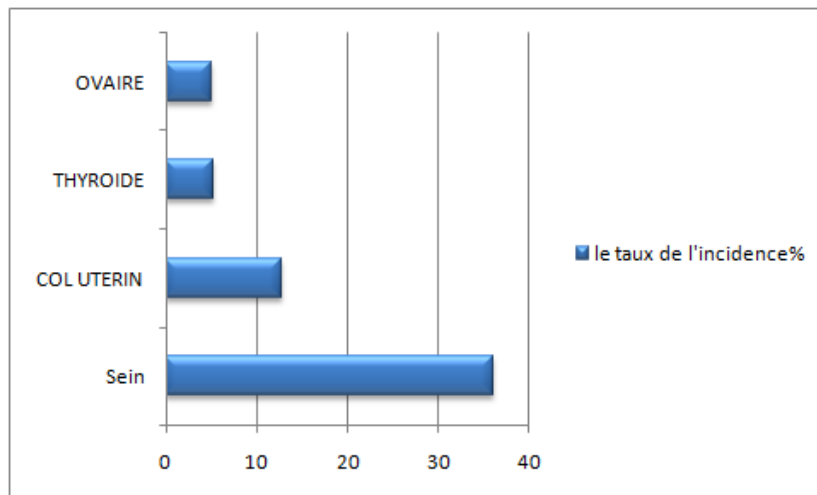


FIGURE 1.1 – Les 4 cancers les plus fréquents chez la femme.

Dans la plupart des pays, un grand effort est fait aujourd'hui pour obtenir des statis-

tiques fiables, afin d'en savoir plus sur les causes et les facteurs de risques du cancer. Telles statistiques sont présentées dans le tableau (1.1).

Tableau 1.1 – Répartition du nombre de cas des principaux types de cancer enregistrés à l'Institut National d'Oncologie (INO) de Rabat (2002 à 2007) [5]

Année	Patients de L'INO				
	Nombre de cas	sein	col	poumon	colon rectum
2002	4896	974	630	297	248
2003	4993	975	601	321	281
2004	5099	1100	617	343	247
2005	4944	1031	576	330	266
2006	5310	1106	749	315	313
2007	5210	1083	660	329	361
Total	30452	6269	3833	1935	1643

Ces statistiques montrent une évolution malgré légère, du cancer du sein de 2002 à 2007 du 19% à 20,78% durant toute cette période, le cancer du sein se montrent le plus important en enregistrant des taux élevés par rapport aux autres cancers diagnostiqués qui ne dépassent pas 12% pour le cancer du col de l'utérus du 6% pour le cancer du poumon. Ces chiffres montrent la nécessité d'un diagnostic précoce efficace et fiable.

Donc, le cancer du sein reste le cancer le plus élevé. Malgré la connaissance des mécanismes de développement des cancers, les causes du cancer du sein ne sont pas connues. Pour cela, le seul moyen est la détection et le diagnostique précoce basés sur l'exploration des seins par la mammographie et l'échographie associées parfois à l'IRM.

## 1. Anatomie du sein

Le sein est un organe essentiellement composé d'une glande mammaire, tissu adipeux, muscle pectoral et canaux lactifères qui acheminent le lait sécrété par les lobules galactophoriques [11]. Le tissu glandulaire, responsable de la production exocrine de lait, est constitué de 15 à 20 lobes composés de nombreux lobules qui sont séparés entre eux par du tissu graisseux et des bandes de tissu conjonctif, appelées ligaments de Cooper. Plusieurs canaux lobulaires forment un canal galactophore qui s'abouche dans un conduit lactifère. L'ensemble des lobules drainés par un conduit lactifère forme un lobe glandulaire. Les conduits lactifères convergent pour s'ouvrir au sommet du mamelon. Le tout est recouvert par la peau. La quantité de ces composantes peut varier d'une femme à l'autre qui influence sur sa forme et son volume. La figure (1.2) détaille ces différents composantes mammaires.

Durant la puberté, le sein se développe grâce à l'hormone oestrogène produite par les ovaires. Les seins sont très sensibles aux changements hormonaux. Pendant la grossesse, d'autres changements hormonaux préparent le sein pour la production du lait. Avec la ménopause, la production d'oestrogène diminue et le tissu glandulaire s'atrophie et est remplacé par du tissu graisseux [1]. La plupart des maladies, qu'elles soient bénignes ou malignes,

prennent origine des lobules et des canaux. Le drainage lymphatique se fait principalement vers les ganglions axillaires qui sont atteints s'il y a propagation du cancer.

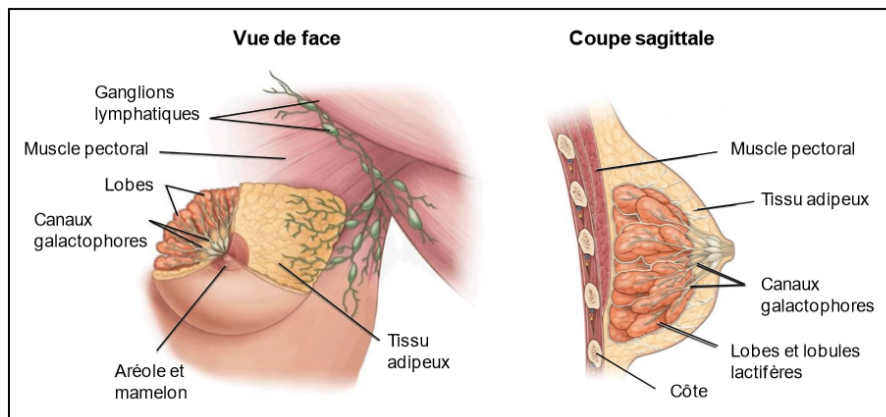


FIGURE 1.2 – Anatomie du sein [1]

## 2. Dépistage

Le but du dépistage du cancer est de détecter, au plus tôt, en l'absence de symptômes, des lésions susceptibles d'être cancéreuses ou lésions précancéreuses. L'intérêt du dépistage est de pouvoir ainsi détecter plus précocement un cancer, de mieux soigner le patient et de réduire le taux de mortalité. Il est aujourd'hui possible de dépister ou de détecter précocement certains cancers. Le dépistage peut être réalisé soit dans le cadre d'un programme organisé par les autorités de santé publique, soit de façon individuelle à l'initiative du professionnel de la santé ou de la patiente [12].

Le dépistage du cancer du sein est caractérisé par trois phases à savoir : un interrogatoire suivi d'un examen clinique et un examen radiologique complété éventuellement par un examen biologique. Le but de l'examen radiologique est de vérifier la présence d'une anomalie radiologique afin de déterminer la conduite adéquate à tenir. Malgré l'existence de plusieurs modalités d'imagerie médicale, la mammographie, qui est une radiographie des seins, est l'examen de référence utilisé lors des campagnes de dépistage du cancer du sein en raison de sa capacité à détecter les cancers de petite taille bien avant que les femmes ne soient en mesure de les palper ou que les premiers symptômes ne se manifestent. Ainsi, elle permet de détecter, plus de 90% des cancers. Dans les programmes de dépistage, il est réalisé sur chaque femme quatre mammographies, à raison d'une mammographie de face et de profil pour chaque sein. Bien que cela ne soit pas systématique, il peut arriver que le radiologue demande des examens complémentaires tels que l'échographie ou l'imagerie par résonance magnétique lorsqu'il est en présence de seins denses rendant d'éventuelles lésions plus difficiles à détecter. Si après l'examen des mammographies, le radiologue déclare que la mammographie est normale, les clichés de ces mammographies sont envoyés en seconde lecture, pour qu'un autre expert puisse se prononcer sur les mêmes images, afin de ne pas manquer un signe radiologique suspect. En dehors des images de dépistage, la mammographie diagnostique est réalisée lorsqu'une femme ou son médecin a remarqué au cours

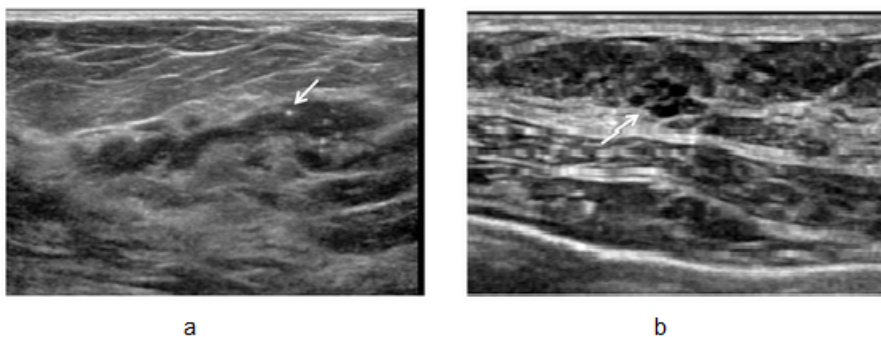
de l'examen du sein une anomalie parmi celles énumérées plus haut. La mammographie diagnostique peut être unilatérale ou bilatérale selon les signes observés ; son but est de rechercher un signe radiologique dans une zone suspecte ou d'analyser de façon plus précise une anomalie détectée cliniquement. Contrairement à la mammographie de dépistage qui se fait à intervalle de deux ans pour chacune des femmes se trouvant entre 50 et 74 ans, il n'y a pas d'intervalle de temps requis pour la mammographie diagnostique. Une mammographie de contrôle peut donc être réalisée quatre à six mois après la mammographie diagnostique en fonction de l'anomalie observée[13].

### 3. Modalités d'imagerie

Dans plusieurs pays du monde, la démarche diagnostique ou dépistage liée au cancer du sein est caractérisée par trois phases à savoir : un interrogatoire suivi d'un examen clinique, un examen radiologique complété éventuellement par un examen biologique. Le but de l'examen radiologique est de vérifier la présence d'une anomalie. Dans le cadre des campagnes de dépistage individuel ou collectif, une ou plusieurs modalités d'imagerie médicale peuvent être utilisées afin de compléter l'examen clinique : l'échographie (imagerie par ultrasons), l'IRM (Imagerie par Résonance Magnétique) et la mammographie (imagerie par rayons X). Dans ce qui suit, nous présentons le principe de fonctionnement de ces différentes techniques [11].

#### 3.1. L'échographie

L'échographie est une méthode d'imagerie médicale qui consiste à appliquer sur le sein une sonde qui envoie des ultrasons qui traversent les tissus puis sont renvoyés sous la forme d'un signal qui s'affiche comme une image sur l'écran. Cette technique présente des avantages et des inconvénients. Parmi les avantages de l'échographie : elle est peu coûteuse et ne présente aucun risque pour la patiente et utilisée même dans le cas où la patiente est enceinte. Ainsi, l'échographie mammaire est un examen très utile pour évaluer la nature liquide ou solide des masses palpées ou découvertes sur les mammographies, et elle peut aider à localiser l'anomalie pour faciliter un prélèvement ou reconnaître s'il s'agit d'un kyste liquidien, mais elle ne peut jamais remplacer la mammographie car l'échographie n'est pas un examen de dépistage du cancer du sein de première intention [14]. A cela s'ajoute l'aspect temps-réel de l'échographie qui permet à la fois de comparer ce que l'on sent avec les doigts au cours de la palpation à ce que l'on observe sur l'écran et de guider le radiologue lorsqu'il réalise un prélèvement. De plus, l'échographie est indispensable lorsque la densité des seins ne permet pas d'avoir une mammographie de bonne qualité [13]. Pour les inconvénients, on peut citer la difficulté de détecter les objets de petites tailles comme les microcalcifications et les petites masses ainsi que le bruit induit par les ultrasons et qui altère la qualité des images échographiques telles qu'il est montré sur la figure (1.3). Toutes les anomalies détectées par une échographie du sein ne sont pas cancéreuses. Selon l'anomalie détectée et son apparence suspecte, le médecin décide ou non de réaliser d'autres examens comme une biopsie qui lui permettront de confirmer ou non le diagnostic de cancer.



---

FIGURE 1.3 – Echographie présentant : a) des microcalcifications, b) des microkystes

### 3.2. La mammographie

Les mammographies sont des examens radiographiques, utilisant les rayons X pour analyser les tissus intérieurs du sein et détecter au plus tôt des anomalies dans la glande mammaire avant même qu'elles n'aient provoquées des symptômes cliniques. Pour obtenir des images de bonne qualité, le radiologue doit faire une analyse précise. Chaque sein est comprimé entre le détecteur du mammographe et une plaque transparente de compression pour bien étaler le tissu du sein et pour éviter la superposition des structures du sein, et elle limitera aussi le mouvement qui pourrait brouiller l'image radiographique. Le but de la mammographie est de produire de façon systématique des images de haute résolution pour une appréciation plus détaillée des microcalcifications. Cet examen est indiqué : chez toutes les femmes après 40 ans et doit être réalisé tous les deux à trois ans, ou lors de la découverte d'une boule dans le sein, d'un écoulement anormal du mamelon ou d'une anomalie de la peau du sein [12]. Cette technique présente des avantages et des inconvénients. Parmi les avantages de la mammographie : elle est peu coûteuse et permet de détecter les objets de petite taille. Pour les inconvénients : elle contient des bruits. La figure (1.4) représente une image mammographique du sein [11].

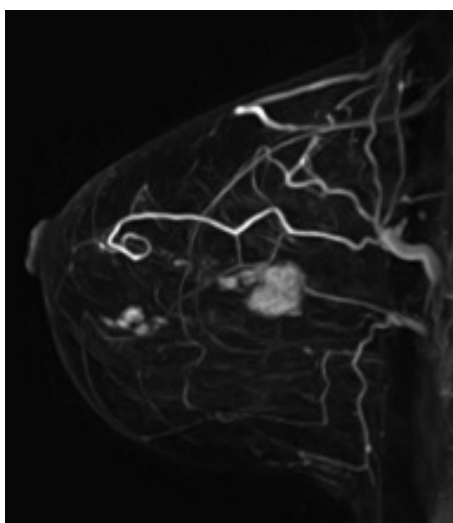


---

FIGURE 1.4 – Image mammographique d'un sein.

### 3.3. L'imagerie par résonance magnétique

L'Imagerie par Résonance Magnétique (IRM) est une technique récente d'imagerie médicale. Apparue au milieu des années 1940, basée sur des phénomènes physiques complexes. Elle est très vite devenue, en médecine une technique d'étude essentielle de la moelle épinière, du rachis et du l'encéphale puis, plus récemment de l'os et des articulations; des indications nouvelles apparaissent avec l'étude du coeur et des vaisseaux, du pelvis, du foie, du sein, ect[15]. L'IRM est une technique d'imagerie médicale complémentaire pour préciser la présence des anomalies du sein chez les femmes à risque diagnostiquées par la mammographie et/ou à l'échographie. Cet examen montre parfois mieux l'extension précise du cancer ou la présence de cancers additionnels. L'IRM utilise des ondes de radiofréquences qui consiste à faire vibrer des atômes d'hydrogène du corps humain, placés dans un champ magnétique puissant. Tous les atômes d'hydrogène s'orientent vers la même direction. On dit qu'ils sont mis en résonance. À l'arrêt de cette stimulation, les atômes restituent l'énergie accumulée en produisant un signal. Ce dernier est enregistré et traité sous forme d'image par un système informatique. Cette technique présente des avantages et des inconvénients. Parmi les avantages de l'IRM : elle réservée à des cas difficiles, permet de détecter les objets de petite tailles. De plus, les rayonnements sont totalement inoffensifs. Cet examen fournit des informations multi-paramétriques. Pour les inconvénients : elle contient des bruits et plus coûteuse. L'inconfort du patient durant cet examen : le sujet est maintenu à l'intérieur d'un tube étroit et subit un bruit fort. Le bruit produit par les appareils fournissant à cadence élevée des impulsions électromagnétiques et aussi pour des contre-indications. Vous devez impérativement prévenir le radiologue dans les cas suivants : vous êtes porteur d'un Pace-Maker (stimulateur cardiaque) ou de prothèses valvulaires (valves cardiaques), vous êtes porteur de prothèses métalliques (plaques de colonne vertébrale, prothèse de hanche, clips métalliques), vous avez subi une intervention chirurgicale au niveau du cerveau ou de l'oeil, vous êtes porteur d'implants auditifs. En fonction des renseignements que vous lui fournirez, le radiologue décidera de poursuivre l'examen ou de vous proposer une alternative vers d'autres techniques d'imagerie. La figure (1.5) représente une image IRM du sein [11].



---

FIGURE 1.5 – IRM d'un sein



## 4. Les tumeurs du sein

Le mot tumeur est un terme générique correspondant au développement d'un tissu nouvellement formé au sein d'un tissu normal. Elle est provoquée par le dysfonctionnement du développement cellulaire. Les cellules malignes se multiplient sans contrôle jusqu'à créer une tumeur qui s'attaque aux tissus sains avoisinants. Cette tumeur peut propager des cellules cancéreuses dans tout l'organisme : on dit alors qu'elle devient métastatique [16]. Contrairement aux idées reçues, le terme tumeur est utilisé non seulement pour des pathologies cancéreuses, mais aussi pour d'autres productions à caractère bénigne. La bénignité (nature non cancéreuse) ou la malignité (cancer) peuvent être évoquées selon l'aspect retrouvé sur les examens d'imagerie pratiqués, mais seront souvent confirmées par la réalisation d'une biopsie [17].

### 4.1. Les tumeurs mammaires non cancéreuses

Les tumeurs bénignes ont des contours bien délimités. Elles ont une croissance lente et restent localisées dans le tissu dans lequel elles sont apparues. Les tumeurs bénignes sont composées de cellules qui ressemblent aux cellules normales du tissu concerné. La tumeur bénigne qui se développe le plus souvent dans le sein est le fibroadénome. Les autres affections bénignes du sein sont [18] :

- **Le kyste** : le kyste du sein est un sac rempli de liquide qui se développe dans le sein. Il s'agit du type le plus courant de masse bénigne au sein chez les femmes de 35 à 50 ans. Un kyste mammaire est habituellement bénin et rarement cancéreux. De plus, le fait d'avoir un kyste dans un sein n'accroît pas le risque de cancer du sein.

- **Les changements fibrokystiques** : les changements fibrokystiques du sein, aussi appelés mastose sclérokystique ou fibrose kystique du sein, sont des affections bénignes courantes qui se manifestent chez 50 à 80 % des femmes. Ces changements apparaissent souvent dans les deux seins. Ils ne font pas augmenter le risque de cancer du sein chez la plupart des femmes. Cependant, la présence de changements fibrokystiques du sein risque d'accroître légèrement la possibilité de développer un cancer du sein s'il y a des antécédents familiaux de cette maladie chez une parente du premier degré (mère, sœur, fille).

- **L'hyperplasie** : l'hyperplasie est une affection bénigne du sein. Une hyperplasie signifie la prolifération de cellules, c'est-à-dire que le nombre de cellules qui tapissent les canaux ou les lobules du sein augmente.

- **L'écoulement du mamelon** : l'écoulement mamelonnaire correspond à une fuite de liquide d'un ou des deux mamelons. Il se produit habituellement après un accouchement puisque les seins produisent du lait.

- **La gynécomastie** : la gynécomastie est une affection bénigne du sein courante chez l'homme. Les personnes atteintes de cette affection ont des seins plus gros que la normale. Le tissu mammaire est ferme. Il est habituellement constitué de tissu canalaire et de tissu conjonctif. La gynécomastie peut rendre plus difficile le diagnostic de cancer du sein.

La plupart des masses du sein ne sont pas synonymes de cancer, mais seul l'examen anatomopathologique réalisé après une biopsie permet de vérifier qu'il s'agit ou pas d'un cancer.

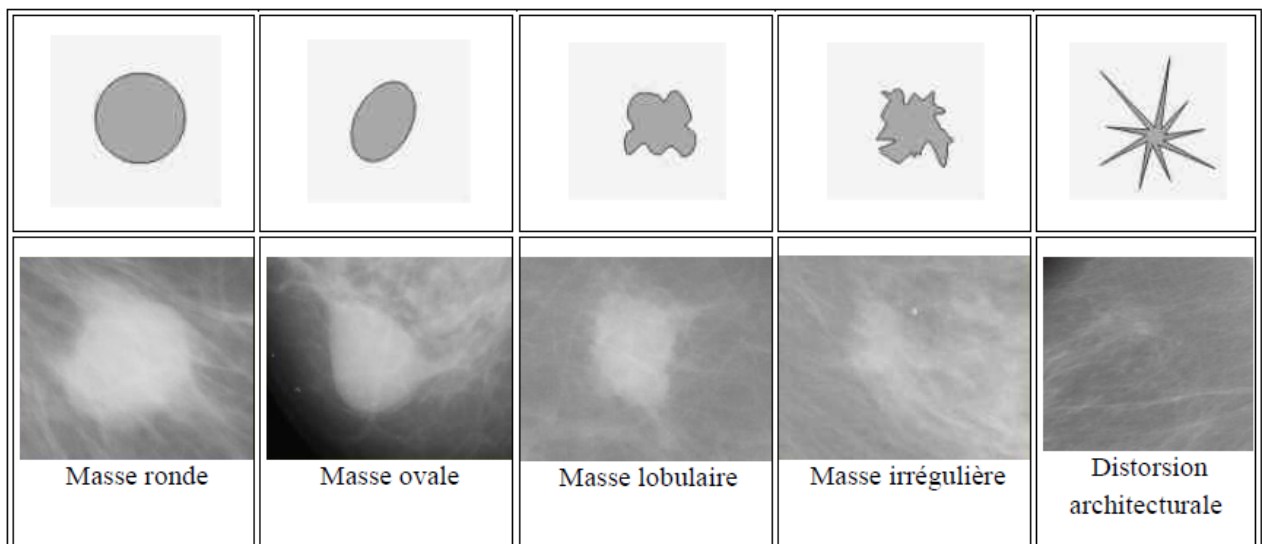
## 4.2. Les tumeurs mammaires cancéreuses

Les tumeurs malignes ou les cellules cancéreuses ont des contours le plus souvent mal délimités et ont tendance à envahir les tissus voisins pour y former une nouvelle tumeur qu'on appelle métastase. Les cellules cancéreuses sont des cellules qui ont perdu leurs caractéristiques d'origine, elles présentent diverses anomalies par rapport à des cellules normales : (forme et taille différentes, contours irréguliers).

## 5. Les pathologies mammaires

### 5.1. Les masses

Une opacité ou une masse est une lésion importante étant plus dense que le tissu mammaire adjacent. Elles apparaissent blanches sur les mammogrammes. Les différentes caractéristiques de ces masses sont à savoir la forme, le contour et la densité qui sont utiles pour le radiologue dans l'évaluation de la probabilité d'avoir cancer. La plupart des masses bénignes sont bien circonscrites, compact, et à peu près circulaire ou elliptique, comme le montre la figure (1.6). Les lésions malignes ont généralement une limite floue, aspect irrégulier et parfois sont entourés par un motif de rayonnement de spicules linéaires [19].




---

FIGURE 1.6 – Formes des masses.

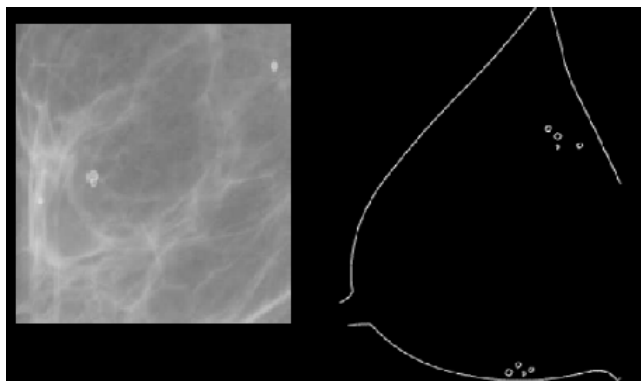
## 5.2. Les microcalcifications

Les calcifications sont des petits minéraux (calcium) déposés dans le tissu mammaire qui apparaissent comme des régions à haute intensité localisées dans la mammographie : elles apparaissent comme des ponctuations blanches sur les tissus. Il existe deux types de calcifications : microcalcifications et macrocalcifications. Les macrocalcifications sont les dépôts de calcium grossiers, dispersés. Ces dépôts sont généralement associées à des affections bénignes et nécessitent rarement une biopsie du sein. Les microcalcifications sont des dépôts de calcium isolés qui apparaissent normalement en grappes ou se trouvent noyées dans une lésion. Les microcalcifications individuels varient généralement de taille de 0,1 à 1,0 mm avec un diamètre moyen d'environ 0,5 mm [20, 21].

### 5.2.1. L'étude morphologique des microcalcifications

L'objectif de l'étude de la morphologie des microcalcifications a pour but de séparer les microcalcifications bénignes des malignes. Les microcalcifications bénignes sont des microcalcifications arrondies ou ovales, uniformes dans leur taille et leur forme. À l'inverse, celles qui sont irrégulières sont souvent malignes [22, 23].

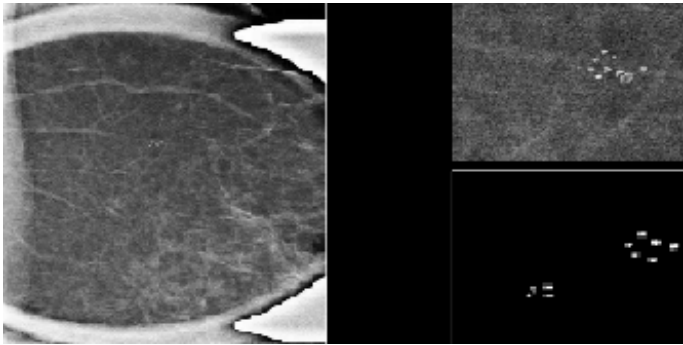
- Microcalcifications cutanées ou dermiques : elles présentent typiquement un centre clair. Des clichés en incidence tangentielle sont souvent utilisés pour confirmer la localisation cutanée de ces microcalcifications (voir figure 1.7).



---

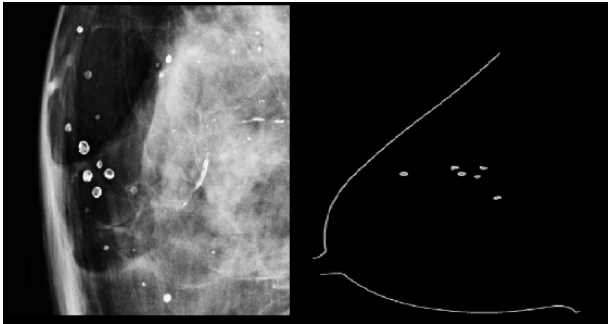
FIGURE 1.7 – Microcalcifications cutanées

- Microcalcifications rondes : elles ont une forme ronde et peuvent être de tailles variables. Lorsqu'elles mesurent moins de 0.5mm, elles sont dites punctiformes ou pulvérulentes (voir figure 1.8).
- Microcalcifications à centres clairs : leur taille peut s'étendre de 1 mm à plus de 1 cm. Elles sont rondes ou ovales, à surface lisse et à centre clair. La paroi calcifiée qui les entoure est plus épaisse que celle des microcalcifications en coquille d'oeuf (voir figure 1.9).
- Microcalcifications dystrophiques : elles mesurent habituellement plus de 0.5 mm de diamètre et sont de formes irrégulières. Elles présentent parfois un centre clair. Ces microcalcifications sont souvent rencontrées dans un sein irradié ou après un traumatisme mammaire (voir figure 1.10).



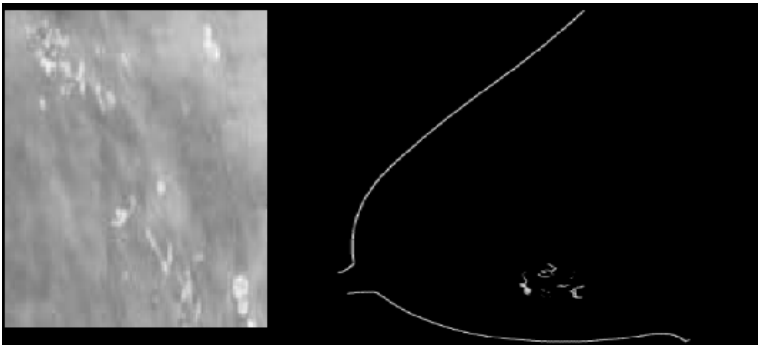
---

FIGURE 1.8 – Microcalcifications rondes



---

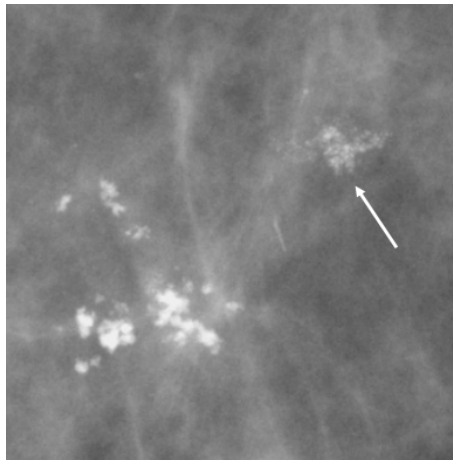
FIGURE 1.9 – Microcalcifications à centres clairs



---

FIGURE 1.10 – Microcalcifications dystrophiques

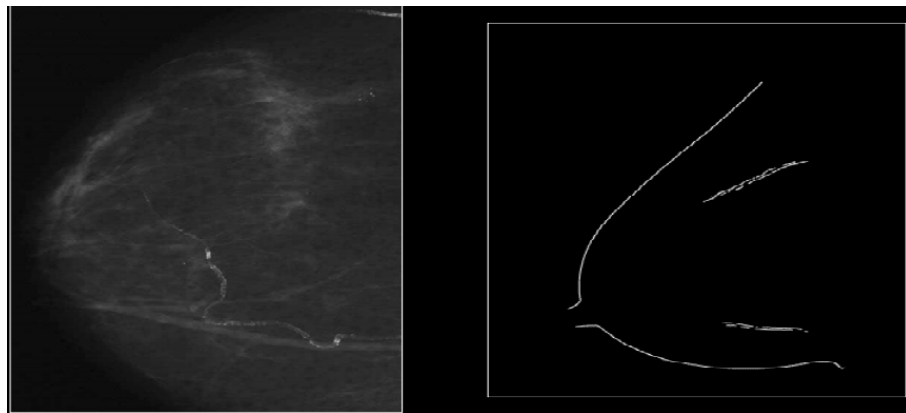
- Microcalcifications amorphes ou indistinctes : elles sont souvent plus ou moins rondes ou en forme de flocons. Elles sont de petites tailles et généralement à contours vagues sans forme spécifique (voir figure 1.11).



---

FIGURE 1.11 – Microcalcifications amorphes

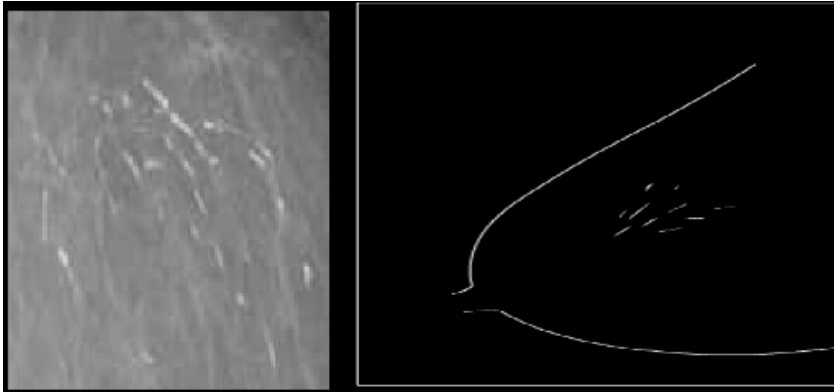
- Microcalcifications vasculaires : ces microcalcifications en rails ou linéaires (voir figure 1.12).



---

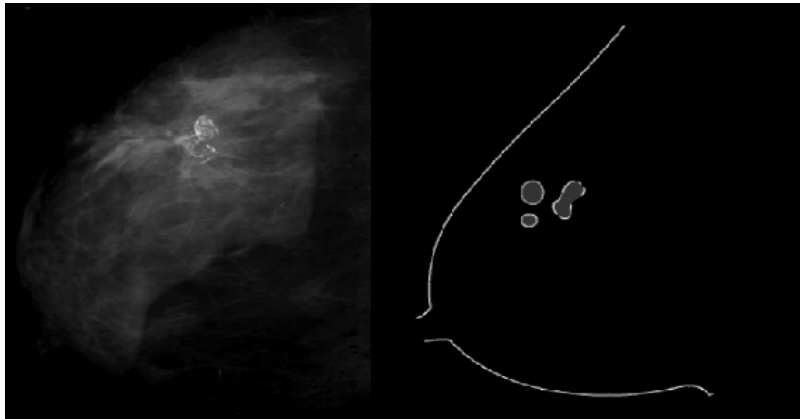
FIGURE 1.12 – Microcalcifications vasculaires

- Microcalcifications en bâtonnets : : elles sont généralement associées à une ectasie canalaire (dilatation du canal galactophore) et sont alors dirigées vers le mamelon. Elles sont souvent retrouvées chez les patientes de plus de 60 ans (voir figure 1.13).
- Microcalcifications d’oeuf : ces microcalcifications très fines apparaissent comme des dépôts calciques sur la surface d’une sphère. Ce dépôt mesure généralement moins de 1 mm (voir figure 1.14).
- Microcalcifications de suture : elles correspondent à des dépôts calciques sur du matériel de suture. Ces microcalcifications sont typiquement linéaires ou tubulaires et présentent fréquemment des noeuds (voir figure 1.15).



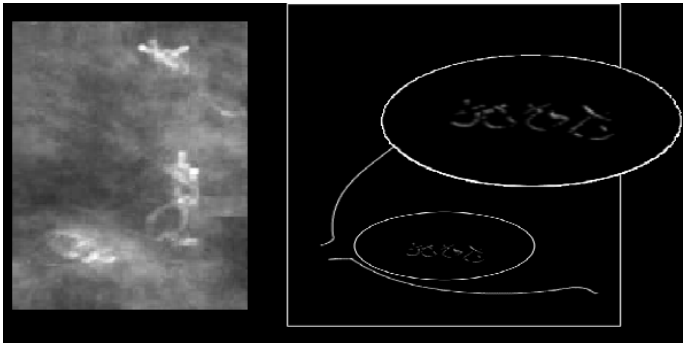
---

FIGURE 1.13 – Microcalcifications en bâtonnets



---

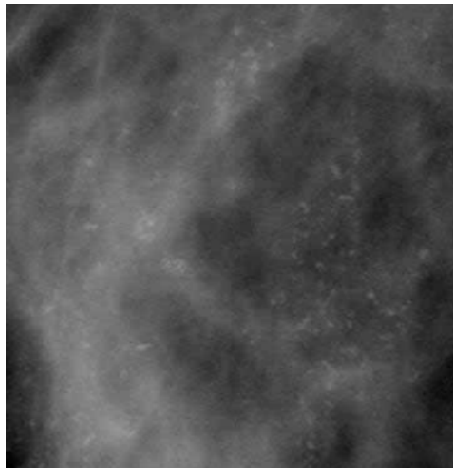
FIGURE 1.14 – Microcalcifications d’oeuf



---

FIGURE 1.15 – Microcalcifications de suture

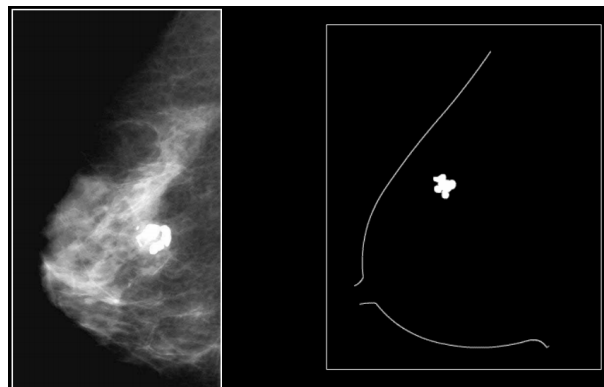
- Microcalcifications fines et polymorphes : elles sont habituellement mieux visibles que les microcalcifications amorphes. Elles sont de taille irrégulières et de forme variables mesurant généralement moins de 0.5 mm de diamètre (voir figure 1.16).



---

FIGURE 1.16 – Microcalcifications polymorphes

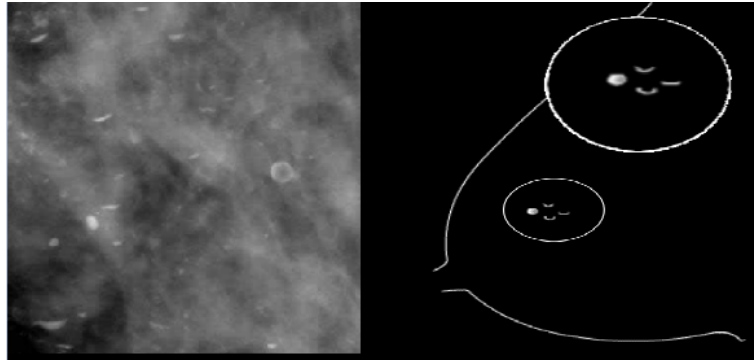
- Microcalcifications grossières ou coralliformes : elles sont de grande taille supérieures à 2/3 mm de diamètre (voir figure 1.17).



---

FIGURE 1.17 – Microcalcifications grossières

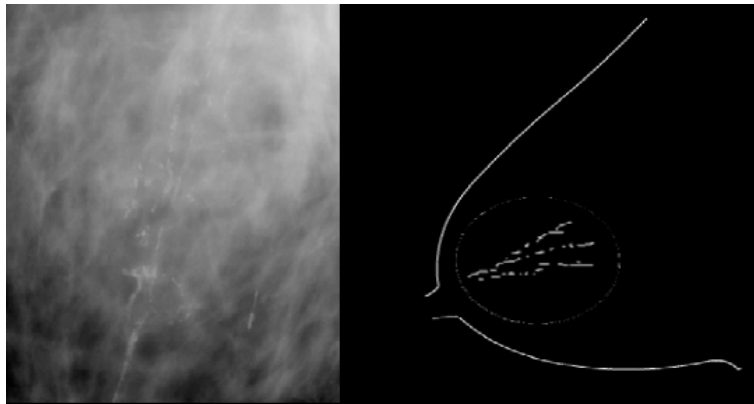
- Microcalcifications à type de lait calcique : elles sont nettes, semilunaires, en croissant ou curvilignes à concavité supérieure, ou encore linéaire dessinant la partie déclive des kystes (voir figure 1.18).




---

FIGURE 1.18 – Microcalcifications à type de lait calcique

- Microcalcifications linéaires et ramifiées : elles mesurent moins de 0.5 mm d'épaisseur. Elles sont irrégulières et de formes parfois linéaires ou curvilignes généralement discontinues, coudées ou branchées (voir figure 1.19).




---

FIGURE 1.19 – Microcalcifications linéaires

### 5.2.2. Distribution

La distribution des microcalcifications est un critère fondamental. Elle présente leur répartition dans le sein et joue un rôle important dans la prise de décision de la malignité [23]. Les différentes distributions possibles des microcalcifications sont représentées sur la figure (1.20)

*a) Microcalcifications diffuses/éparses* : elles sont distribuées aléatoirement dans le sein.

*b) Distribution régionale* : les microcalcifications sont dispersées dans un large volume du tissu mammaire et ne présentent pas une distribution canalaire.

*c) Microcalcifications groupées, en amas ou en cluster* : ces termes sont utilisés lorsque de multiples microcalcifications (au moins cinq) occupent un petit volume tissulaire.



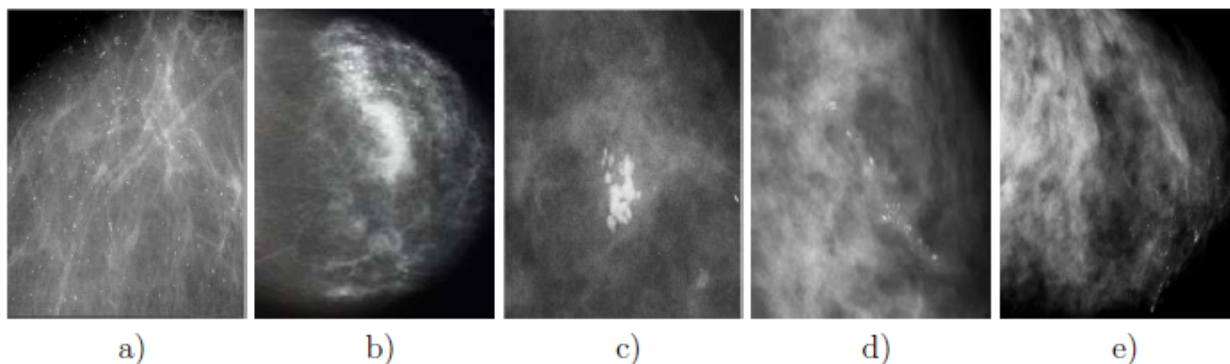


FIGURE 1.20 – Les différentes distributions des microcalcifications : a) Mcs diffuses ou éparses : b) Régionale, c) Cluster, d) Linéaire et e) Segmentaire.

*d) Distribution linéaire* : les microcalcifications sont disposées les unes derrière les autres sous forme d'une ligne. Il s'agit généralement de dépôts calciques dans un galactophore.

*e) 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.

## 6. Symptômes

On appelle symptôme d'une maladie, toute manifestation anormale provoquée par cette maladie. Les symptômes listés ci-dessous ne signifient pas nécessairement qu'il s'agit d'un cancer du sein. Mais si c'est le cas, Ils doivent donc être évalués par un médecin. Il ne faut pas attendre et ne négliger aucun signe inhabituel et voici les principaux [24] :

- Petit amas dans le sein, douloureux ou non.
- Petit amas ou enflure à l'aisselle.
- Bosse au sein, qu'elle soit fixe ou mobile. Il s'agit du symptôme le plus fréquent, pour les femmes et les hommes.
- Modification de la taille, de la texture ou de la forme des seins.
- Changement d'apparence de la peau du sein : épaissement ou durcissement, rougeur inhabituelle, chaleur, changement de la peau autour du mamelon.
- rétraction du mamelon (le mamelon est retourné vers l'intérieur).

## 7. Facteurs de risque

Malgré les progrès qui ont permis de mieux connaître les mécanismes de développement des cancers, les causes du cancer du sein ne sont actuellement pas connues. Néanmoins, les études ont mis en évidence certains facteurs de risque qui favorisent le développement du cancer du sein. Il y a différents types de facteurs de risque [25] :

- L'âge compris entre 45 et 75 ans : comme pour la plupart des maladies chroniques, le risque d'avoir un cancer du sein augmente avec l'âge [26].
- Age aux premières règles, en ce qui concerne la population générale, de nombreuses études montrent que la survenue des premières règles avant l'âge de 12 ans augmente le risque de cancer du sein [27, 28].
- Les facteurs familiaux ou génétiques : la présence de cas de cancer du sein dans la famille particulièrement du côté maternel (grand-mère, mère, tante ou soeur)[29, 30].
- L'origine géographique influence aussi le risque de cancer du sein. Ce risque augmente de plus de 4 fois si la patiente est née en Amérique du Nord ou en Europe du Nord comparé à une femme née en Asie ou en Afrique [26].
- Les caractéristiques propres à l'individu : le risque hormonal (puberté précoce [31] , ménopause tardive [32], l'âge tardif de la première grossesse etc), la consommation exagérée d'alcool, la consommation de tabac [33, 34].
- Les traitements hormonaux substitutifs de la ménopause sont prescrits pour pallier la diminution du niveau d'hormones ovariennes circulantes [35, 36], et le risque de cancer du sein augmente avec la durée d'utilisation.
- L'allaitement et la durée d'allaitement sont associés à une diminution du risque de cancer du sein, une étude internationale effectuée à partir de 47 études, a rapporté une diminution significative du risque de cancer du sein de plus de 4% pour chaque période d'allaitement de 12 mois [37]. De plus, l'allaitement retarde le retour des menstruations et de la production hormonale ovarienne , en particulier une réduction d'oestrogènes. Le cancer du sein est un cancer hormono-dépendant. Par conséquent, la lactation réprimerait l'apparition et le développement du cancer du sein [38].

## 8. Base de données mammographies

Pour évaluer les performances d'une nouvelle approche d'analyse, il est recommandé de les valider sur des bases de données de référence. En effet les chercheurs ont besoin de comparer leurs résultats sur les mêmes images. Malheureusement, la plupart des bases de données d'images médicales réelles ne sont pas toujours accessibles au public cela est dû aux issues de la vie privée du patient. Actuellement, les bases de mammogrammes numériques les plus utilisées par les chercheurs sont MIAS et DDSM.

## 8.1. Base de données MIAS

La base de données MIAS (Mammography Image Analysis Society) est une base de données de mammographie numérique. Les films tirés par le programme national de dépistage du cancer au Royaume-Uni ont été numérisés à 50 microns de pixels avec un microdensitomètre Joyce-Loebl balayage, un dispositif linéaire dans la gamme de 0 à 3,2 et la densité optique représentant chaque pixel avec un mot de 8 bits. La base de données contient 322 films numérisés et sont disponibles sur 8mm (Exabyte) 2,3 Go de bande. Les images sont de 1024x1024 pixels. La base de données MIAS est disponible sur le site web<sup>1</sup>.

Un fichier joint à la base de données décrit en détail les images. Il est organisé en sept colonnes de la manière suivante :

1<sup>ère</sup> colonne : le numéro de référence de l'image dans la base de données MIAS

2<sup>ème</sup> colonne : caractérise le type des tissus du sein. Il y a trois types :

- F : Graisseux (Fatty)
- G : Glandes Graisseux (Fatty-glandular)
- D :Dense (Dense-glandular)

3<sup>ème</sup> colonne : Type de pathologie

- CALC : Calcification (Calcification)
- CIRC : Well-defined/circumscribed masses
- SPIC :Spiculated masses
- MISC :Other, ill-defined masses
- ARCH :Architectural distortion
- ASYM :Asymmetry
- NORM :Normal

4<sup>ème</sup> colonne : Sévérité de pathologie

- B Benin
- M Malin

5<sup>ème</sup> et 6<sup>ème</sup> colonne : les coordonnées (x,y) du centre de la pathologie.

7<sup>ème</sup> colonne : Rayon approximatif (en Pixel) d'un cercle enfermant l'anomalie.

## 8.2. Base de données DDSM

La base de données numérique de dépistage par mammographie (DDSM : Digital Database for Screening Mammography) Marathon de l'université de la Floride du Sud. La base de données MIAS est disponible sur le site web<sup>2</sup>. Une description de cette base a été effectuée par American college of Radiology. La base correspondante contient 2620 dossiers de patientes classées en trois groupes : normaux, bénins et malins. Chaque cas contient les vues obliques externe (MLO) et face externe (CC) de chaque sein. En plus, la base fournit

---

1. <http://www.mammoimage.org/databases>.

2. <http://marathon.csee.usf.edu/Mammography/database.html>

un fichier contenant des informations sur les images comme le scanner, la résolution spatiale et des informations sur la patiente comme l'âge au moment de l'étude, la densité du sein, etc. Pour les images qui contiennent une ou plusieurs anomalies, une marque a été introduite par un radiologue expérimenté localisant les régions contenant les pathologies.

## Conclusion

Dans ce chapitre, nous avons présenté les notions de base concernant l'anatomie et le cancer du sein ainsi que les outils de l'imagerie médicale permettant le diagnostic. Une grande attention a été consacrée à l'étude des spécifications des pathologies mammaires à savoir les masses et les microcalcifications, car l'étude des concepts médicaux s'avère très importante avant de débiter un autre traitement. À noter que le travail du radiologue et de l'informaticien sont très complémentaires dans le domaine de l'imagerie médicale. Nous avons également décrit deux bases de données d'images mammographiques (MIAS et DDMS) utilisées comme bases de référence par les chercheurs. Nous avons choisi de travailler avec la base MIAS pour la simplicité d'exploitation par rapport à la base DDMS. Le chapitre suivant sera consacré à la présentation des différents travaux de recherche trouvés dans la littérature pour la segmentation et classification des images à niveau de gris dans le cas d'un cancer du sein.

# Chapitre 2

## Technique de segmentation : application aux images mammographiques

### Introduction

Dans ce chapitre, nous présentons un état de l'art sur les différentes phases de traitement des images mammographiques à savoir : la phase de prétraitement, la phase de segmentation des lésions mammaires et la phase de classification. Il permet une meilleure compréhension des choix des méthodes de prétraitement et de segmentation adoptées par la suite. Enfin, l'étude des méthodes de classification permet de sélectionner le classifieur le plus approprié à notre domaine d'application. La figure suivante illustre la procédure de diagnostic assistée par ordinateur :

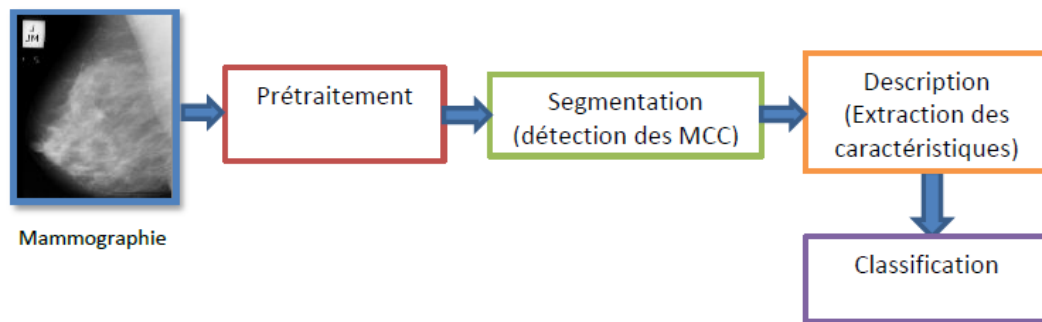


FIGURE 2.1 – Procédure de diagnostic assistée par ordinateur

### 1. Méthodes de prétraitement

Le prétraitement est une première étape qui consiste à améliorer les caractéristiques de l'image de manière à faciliter la visualisation de la structure que l'on recherche à détecter. D'où, les objectifs principaux de cette étape sont la réduction des facteurs de bruit et

l'augmentation du contraste entre la lésion mammaire et le reste de l'image. Le prétraitement des images mammographiques est connu sous le nom d'amélioration du contraste ou de rehaussement.

Parmi les problèmes du rehaussement du contraste on trouve que certaines régions peuvent ne pas être rehaussées et que d'autres régions peuvent être l'objet d'un excès de rehaussement. Un manque de rehaussement du contraste peut causer des Faux Négatifs (FN). En effet, plusieurs détails concernant la lésion peuvent être négligés. et un excès du contraste peut causer des Faux Positifs (FP). Dans ce cas, plusieurs détails inexistantes réellement peuvent s'ajouter à la lésion. Ce qui ne répond pas à l'objectif principal de la détection précoce d'un cancer.

Parmi les approches couramment utilisées en traitement d'images pour effectuer ce type d'opérations on trouve une approche basée sur la modification globale de l'histogramme. Cette méthode consiste à réaffecter les valeurs d'intensité des pixels afin de rendre la nouvelle répartition des intensités plus uniforme. Ceci peut être réalisé par l'égalisation d'histogramme ou par l'étirement d'histogramme Pizer et al, [39]. Toutefois, ces transformations ont l'inconvénient de faire ressortir le bruit dans des images bruitées. Sivaramakrishna et al[40] ont comparé la performance de quatre algorithmes de rehaussement du contraste dans les deux cas d'images mammographiques contenant des microcalcifications et contenant des masses. Les algorithmes comparés sont : l'amélioration du contraste basée sur le voisinage adaptatif ANCE Morrow et al[41], l'égalisation d'histogramme est basée sur le contraste adaptatif limité [39](Pizer et al), Ji et al utilisent le masque flou adaptatif [42] et Laine et al et Pfisterer utilisent les ondelettes [43, 44]. pour atteindre ce but Hassanien et Ali [45] proposent d'utiliser les ensembles flous et l'analyse de l'histogramme de l'image. D'autres formalismes sont aussi utilisés pour accomplir la tâche d'amélioration du contraste par des filtres hautes fréquences [46, 47, 48]. D'autres méthodes ont aussi été proposées comme l'utilisation du masque flou. Jose, K. Sankar, K. Nirmala [49, 50] proposent une nouvelle approche pour la détection des microcalcifications en mammographie numérique en utilisant Nonsampled Contourlet Transform (NSCT).

Ainsi, le prétraitement est une étape essentielle pour améliorer la qualité visuelle de l'image qui doit utiliser un traitement appelé filtrage qui consiste à modifier la distribution fréquentielle des composantes d'une image.

## 1.1. Les différents types de filtres

### 1.1.1. Filtre passe-bas

C'est un filtre de lissage qui permet de supprimer le bruit grâce à la localisation du bruit dans les hautes fréquences et n'affecte pas les composantes de basse fréquence comme les grandes zones uniformes ; car la fréquence se calcule en fonction de la vitesse de changement de caractéristique dans l'image ; ainsi une zone uniforme aura une fréquence faible. Malgré l'avantage de ce filtre, il ya des inconvénients comme le fait de rendre l'image flou ; ce qui se traduit par une perte des petits détails et des contours.

### 1.1.2. Filtre passe-haut

C'est un filtre d'accentuation à des caractéristiques inverses du filtre passe-bas. Ce filtre est utilisé pour traiter les petits détails d'une image correspondant à des hautes fréquences, mais doit atténuer les composantes de basse fréquence.

— Filtre Homomorphique

En général, une image peut être considérée comme une fonction à deux dimensions de la forme  $I(x, y)$ . Nous pouvons dire que l'image est générée à partir d'un processus physique, ses valeurs sont proportionnelles à l'énergie rayonnée par une source physique. En d'autres termes, une image est une fonction constituée du produit de l'illumination (la quantité de lumière incidente de la source sur la scène en cours de visualisation) par la réflectance (la quantité de lumière réfléchie par les objets de la scène)[51]. Si l'on note  $L(x, y)$  illumination et  $R(x, y)$  la réflexion, l'image  $I(x, y)$  peut être exprimée comme décrit dans l'équation (2.1).

$$I(x, y) = L(x, y).R(x, y) \quad (2.1)$$

— La base de Filtre Homomorphique

Comme déjà indiqué dans la section précédente, les résultats d'illumination des conditions d'éclairage présent lorsque l'image est capturée, peuvent changer lorsque les conditions d'éclairage changent. Nous tenons à améliorer la réflexion tout en réduisant la contribution de l'éclairage, par conséquent, nous avons besoin de séparer les deux composantes de l'équation (2.1), puis on filtre l'image résultante par un filtre homomorphique c'est un passe-haut. Le filtrage homomorphique [52, 53] est un processus de filtrage dans le domaine de fréquence. Si nous pouvions transformer l'équation (2.1) de la multiplication à l'addition, le problème du filtrage passe-haut deviendrait trivial car nous pourrions utiliser la propriété de multiplication ou de convolution de la transformée de Fourier [51].

Pour appliquer la transformée de Fourier nous séparons les des composantes d'illumination  $L(x, y)$  et de réflectivité  $R(x, y)$ .

$$Z(k, l) = \ln(I(x, y)) = \ln(L(x, y)) + \ln(R(x, y)) \quad (2.2)$$

Puis utiliser la transformée de Fourier :

$$TF(Z(k, l)) = TF(\ln(L(x, y))) + TF(\ln(R(x, y))) \quad (2.3)$$

Nous écrivons :

$$Z(u, v) = F_L(u, v) + F_R(u, v) \quad (2.4)$$

Le filtrage de Z par un filtre homomorphique H donne :

$$S(u, v) = H(u, v).Z(u, v) \quad (2.5)$$

$$S(u, v) = H(u, v).F_L(u, v) + H(u, v).F_R(u, v) \quad (2.6)$$

la Transformation de Fourier Inverse de l'équation (2.6) :

$$S(x, y) = ITF(S(u, v)) \quad (2.7)$$

$$S(x, y) = ITF(H(u, v).F_L(u, v)) + ITF(H(u, v).F_R(u, v)) \quad (2.8)$$

L'image traitée est obtenue par l'équation suivante :

$$It(k, l) = \exp(S(x, y)) \quad (2.9)$$

Le filtre passe-haut utilisé dans cette opération est le filtre Butterworth [52] défini comme suit :

$$H(u, v) = \frac{1}{1 + \left(\frac{D(u, v)}{D_0}\right)^{2n}} \quad (2.10)$$

où  $n$  définit l'ordre du filtre.  $D_0$  est la fréquence de coupure et  $D(u, v)$  est donnée par :

$$D(u, v) = \sqrt{u^2 + v^2} \quad (2.11)$$

A la place du filtre classique de Butterworth nous pouvons utiliser un filtre gaussien modifié, l'équation (2.12) utilisé par [54], plus robuste face à la variabilité des conditions d'illumination dans une image.

$$H(u, v) = (r_H - r_L) \cdot (1 - \exp(-(\frac{u^2+v^2}{2\delta_\omega^2}))) + r_L \quad (2.12)$$

Où  $r_H$  et  $r_L$  sont les coefficients maximum et minimum du filtre et  $\delta_\omega$  le facteur fixant la fréquence de coupure.

Lorsque  $r_L < 1$  et  $r_H > 1$  ce filtre atténue les basses fréquences (illumination) et accentue les moyennes et hautes fréquences (réflectance).

### 1.1.3. Filtre médian

Le filtre médian est un filtre non linéaire utilisé pour éliminer le bruit. La réduction de bruit est une étape de pré-traitement classique visant à améliorer les résultats de traitement. La technique de filtre médian est largement utilisée en traitement d'images numériques car il permet, sous certaines conditions, de réduire le bruit tout en conservant les contours de l'image.

Le principe du filtre médian est souvent défini dans le cas d'une image discrète, dont l'implantation pratique est directe [55]. À partir des concepts discrets, il est possible d'en donner une version continue, qui sera adaptée à l'étude théorique de certaines de ses propriétés. Considérons une image discrète  $F$  caractérisée par un niveau de gris  $f(x, y)$ . Soit  $V(x_0, y_0)$  le voisinage associé au point de coordonnées  $(x_0, y_0)$ ; on suppose que ce voisinage comprend  $N$  pixels de coordonnées  $(x_0 - u, y_0 - v)$  avec  $N$  impair.

Soient  $(f_1, f_2, \dots, f_i, \dots, f_{N-1}, f_N)$  les niveaux de gris associés aux  $N$  pixels de  $V(x_0, y_0)$ . Le filtrage médian procède tout d'abord par un tri des valeurs de niveau de gris du voisinage

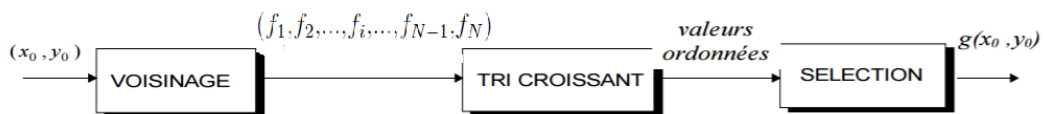


FIGURE 2.2 – Le principe du filtre Médian



suivi d'une sélection de l'élément milieu du tri. Le tri se fait par ordre croissant généralement. Il conduit à former l'ensemble ordonné des valeurs de gris du voisinage de  $f(x_0, y_0)$ . Les éléments ordonnés étant notés  $f_{(i)}$ , le tri croissant est caractérisé par :

$$f_1 < f_2 < \dots < f_{\frac{N+1}{2}} < \dots < f_{N-1} < f_N$$

L'élément médian du voisinage est  $f_{(\frac{N+1}{2})}$ .

Sa propriété est d'être précédé par  $\frac{N-1}{2}$  valeurs inférieures et suivi par autant de valeurs supérieures. Le filtrage consiste à remplacer  $f(x_0, y_0)$  par la valeur médiane du voisinage  $f_{(\frac{N+1}{2})}$ .

Alors l'idée principale du filtre médian est de remplacer chaque entrée par la valeur médiane de son voisinage. Par exemple, si on considère ces neuf pixels :

$$\begin{pmatrix} 5 & 6 & 7 \\ 6 & 200 & 8 \\ 7 & 8 & 9 \end{pmatrix}$$

le filtre médian va considérer les valeurs du voisinage par valeurs croissantes :

$$( 5 \ 6 \ 6 \ 7 \ 7 \ 8 \ 8 \ 9 \ 200 )$$

et prendre la valeur médiane, ici la valeur 7. La valeur du pixel central de la fenêtre est remplacée par la valeur médiane alors à la sortie du filtre donnera :

$$\begin{pmatrix} 5 & 6 & 7 \\ 6 & 7 & 8 \\ 7 & 8 & 9 \end{pmatrix}$$

## 2. Méthodes de segmentation

La segmentation des images consiste à rassembler les pixels entre eux qui partagent un même propriété pour former des régions connexes. Il peut s'agir par exemple de séparer les objets du fond.

### 2.1. Segmentation par seuillage

Le seuillage est la première approche utilisée en segmentation d'image. Il est basé sur le regroupement des pixels semblables. Ce regroupement peut être basé sur une approche globale, locale ou sur une approche hybride. Cette méthode est basée sur les seuils qui sont les minima locaux de l'histogramme alors que les pixels en dessous du seuil représentent l'objet tandis que les pixels au dessus représentent le fond. Cette méthode est efficace lorsque l'histogramme contient des pics séparés [56].

### 2.2. Segmentation par extraction de régions

La segmentation d'images par l'approche régions consiste à découper l'image en régions homogène selon des critères qui peuvent dépendre : de l'intensité, de la couleur, des caractéristiques statistiques ou de la position des pixels [57]. Pour cela, il existe plusieurs techniques dont :

### 2.2.1. Segmentation par croissance de régions

La segmentation par croissance de régions [58], [59], est une méthode qui consiste à faire progressivement grossir les régions autour de leur point de départ qui s'appelle seed. A partir de ces derniers, des régions sont construites par ajout de pixels voisins à la région si l'indicateur d'homogénéité reste vrai [57]. La croissance s'arrête lorsqu'on ne peut plus ajouter de pixel qui ne respecte pas à un critère d'homogénéité.

### 2.2.2. Segmentation par division et fusion

L'idée de cette méthode de segmentation par approche fonctionnant par fusion (Merge) est d'exploiter une image pré-segmentée, ce qui veut dire que l'image contient des régions dont l'homogénéité n'est pas vérifiée. Ces dernières sont regroupées successivement selon un critère donné jusqu'à l'obtention de classes homogènes et la division de région (Split) qui consiste à partitionner l'image en régions homogènes selon un critère donné [60].

## 2.3. Segmentation par détection de contours

Un contour est un ensemble de pixels formant une frontière entre deux ou plusieurs régions voisines. La plupart des algorithmes qui lui sont associés sont locaux ; ils ne considèrent que l'information du pixel. La segmentation se fait en regroupant les pixels en détectant les discontinuités en niveaux de gris, couleurs ou textures dans l'image. Parmi les approches on trouve les méthodes dérivatives qui font intervenir la notion de dérivation afin de détecter les transitions entre les objets et le fond de l'image qui correspondent à des zones de fortes variations d'intensité tels que : Canny, Prewitt, Sobel et Roberts [61, 62, 63, 64]. Une autre méthode basée sur les contours actifs correspondent à des courbes fermées disposées sur une image et faire converger vers une zone particulière d'une image. Ces courbes sont définies dans le domaine des images, elles peuvent se déplacer sous l'influence des forces internes à la courbe et des forces externes calculées à partir des données de l'image [65, 66]. Une autre méthode de segmentation basée sur les structures topographiques d'une image est la ligne de partage des eaux (LPE)[67]. La ligne de partage des eaux est un algorithme introduit en segmentation d'images par Digeabel et Lantuéjoul qui a pour but de diviser une image en termes géographique sous forme des régions homogènes. Elle s'applique au gradient morphologique de l'image originale ou le niveau de gris de chaque point est interprété comme son altitude dans le relief topographique : Un point est d'autant plus élevé dans le relief qu'il est clair dans l'image. Les pixels sombres correspondent donc aux vallées et bassins du relief alors que les pixels clairs correspondent aux collines et lignes de crêtes [68].

— Construction de la LPE

Pour l'obtenir, il faut imaginer l'immersion d'un relief dans l'eau, en précisant que l'eau ne peut pénétrer dans les vallées que par ses minima. La ligne de partage des eaux est représentée par les points où deux lacs disjoints se rejoignent au cours de l'immersion.

— Le processus de LPE par l'inondation de relief :

- 1- La LPE produite par inondation peut-être formulée à l'aide d'une distance topographique [69].
- 2- On prolonge la surface dans un bassin, l'eau va passer par les trous qui s'appellent les minima locaux.
- 3- Quand les eaux issues de deux minima différents se rencontrent on monte une digue

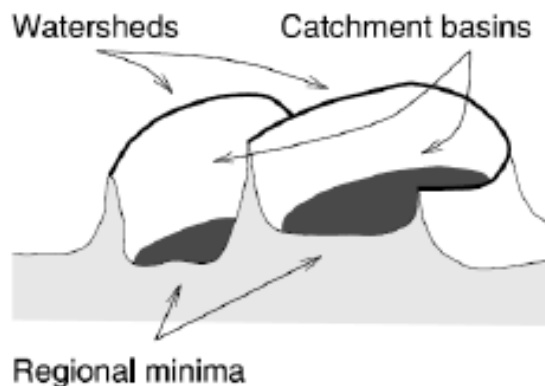


FIGURE 2.3 – Principe de la procédure d’immersion : l’immersion de ses minima régionaux, tout en empêchant la fusion des eaux provenant de différentes sources, partitionne l’image en les bassins versants (catchment bassins) et les lignes de partage des eaux (watersheds)[2]

pour qu’elles ne se mélangent pas.

4- Lorsque la totalité de la surface topographique aura été engloutie, seul les barrages émergeront, délimitant les bassins versants au nombre des minima locaux.

5- A la fin de l’immersion, un ensemble de contour fermé en appelle la ligne de partage des eaux.

La LPE du gradient est sur-segmentée, ceci est dû au fait que les images gradient des images naturelles contiennent un grand nombre de minima dû au bruit ou encore aux textures des régions. Chaque minimum génère un bassin versant dans la LPE.

Il existe plusieurs méthodes qui permettent de pallier à cette sur-segmentation.

La première consiste à filtrer l’image originale afin de supprimer tous les minima non-significatifs : c’est l’approche filtrage.

La deuxième consiste à choisir le nombre de minima locaux et donc le nombre de zones que l’on souhaite mettre en évidence grâce à la LPE : c’est l’approche marqueurs (swamping).

— Les points clés de la segmentation par LPE :

Nous avons vu que les algorithmes de segmentation par ligne de partage des eaux consistent en trois étapes :

1. Extraire des marqueurs des régions à segmenter.
2. Déterminer l’image sur laquelle on calcul la LPE (image gradient le plus souvent).
3. Calculer la LPE associée aux marqueurs.

### 3. Méthodes de détection des microcalcifications

La première classe de signes radiologiques recherchés sont les foyers de microcalcifications. La littérature comporte un nombre important de méthodes de détection : N. Hamdi et al [46] , Wang et al [70], Strickland et al [71], A. Alolfe et al [72], N. Hamad et al [73], Ba-

lakumaran et al [74]. Ils présentent des approches pour détecter les microcalcifications dans la mammographie numérique en se basant sur les ondelettes. Ils décomposent l'image en un ensemble de sous-bandes orthogonales de résolutions différentes. La fréquence la plus basse de sous-bande porte la plupart du temps l'intensité de fond de l'image, et la fréquence la plus haute porte les microcalcifications qui apparaissent dans quelques pixels avec intensité relativement élevée par rapport à leurs pixels voisins. La détection des microcalcifications est obtenue en décomposant le mammographie en différentes sous-bandes de fréquences, en supprimant la sous-bande de basses fréquences et enfin, la reconstruction de la mammographie des sous-bandes ne contenant que les hautes fréquences. Ensuite, un seuil est appliquée pour améliorer la visibilité des microcalcifications.

Cheng et al [75] utilisent une technique basée sur la logique floue pour détecter les microcalcifications. Les structures mammaires non pertinentes sont exclus par un détecteur de courbe. Ensuite, les microcalcifications sont localises à l'aide d'une sélection de seuil itératif et les pixels sont éliminés en employant la morphologie mathématique.

Chan et al [76] utilisent une différence d'image calculé à partir d'un signal d'image, et un signal de suppression d'aide à l'image pour supprimer l'arrière-plan structuré en une mammographie. Ensuite, ils utilisent une technique d'extraction de signal pour détecter des microcalcifications.

T. Stojic et al [77] ont développé un système qui agit sur l'amélioration du contraste de la mammographie pour détecter les (Mcs), en utilisant deux approches : la première méthode est basée sur la théorie multi-fractal, et la seconde est basée sur la morphologie mathématique. Enfin, plus récemment, dans le même objectif S. Al-Kindi et al [78] ont reformulé le problème de l'amélioration de contraste non seulement mammographies mais aussi des images ultrasons pour créer un contraste optimal pour une meilleure visualisation sur les masses et les (Mcs).

Huddin [79] a présenté une autre méthode pour extraire les caractéristiques et de classer les groupes de microcalcifications utilisant la décomposition de la pyramide orientable. Le procédé est motivé par le fait que les amas de microcalcifications peuvent être de tailles et d'orientations arbitraires. Ainsi, il est important d'extraire les caractéristiques dans toutes les orientations possibles pour capturer la plupart des informations de distinction pour la classification.

Eltoukhy [80] a présenté une approche pour le diagnostic du cancer du sein en mammographie numérique en utilisant la transformation en curvelet. Après la décomposition des images de mammographie à base curvelet, une série spéciale de plus grands coefficients est extrait comme vecteur de caractéristique. La distance euclidienne est ensuite utilisé pour construire un classificateur supervisé.

Eltoukhy [81] a aussi présenté une étude comparative entre les ondelettes et transformée en curvelets pour le diagnostic du cancer du sein. L'utilisation de l'analyse multi-résolution sur les images mammographie qui sont décomposés en différents niveaux de résolution, qui sont sensibles à différentes bandes de fréquence.

L. Jasmine [82] a présenté une nouvelle approche pour la détection des microcalcifications en mammographie numérique en utilisant Nonsubsampled Contourlet Transform (NSCT) et Support Vector Machine (SVM). La classification des microcalcifications est obtenu par extraction des caractéristiques de microcalcifications à partir des coefficients Contourlet de l'image et les résultats sont utilisés comme une contribution à la SVM pour la classification. Bhangale [83] a utilisé la banque de filtre de Gabor pour l'analyse de la texture de la mam-

mographie pour détecter les microcalcifications.

Lakshmanan [84] a présenté une approche de classification et détection des microcalcifications dans les images mammographique utilisant la morphologie mathématique et transformation en Contourlet et réseaux de neurones.

### 3.1. Filtre de Gabor

Le filtre de Gabor est une fonction sinusoidale modulée par une enveloppe gaussienne inventé par le physicien anglais Dennis Gabor [85]. La fonction sinusoidale est caractérisée par sa fréquence et par son orientation dont le fonctionnement est proche des traitements visuels humains. Ce sont les filtres les plus utilisés pour décrire la texture d'une image. Ainsi, le filtre de Gabor peut être vu comme un détecteur d'arêtes d'orientation particulières, puisqu'il réagit aux arêtes perpendiculaires à la direction de propagation du sinus [86]. En traitement d'images, on travaille dans le domaine spatial de deux dimensions, ce qui permet de d'écrire la fonction de gabor de la manière suivante [87] :

$$g(x, y) = \exp(-0.5 * ((\frac{\hat{x}}{\sigma_x})^2 + (\frac{\hat{y}}{\sigma_y})^2)) * \cos(2 * \Pi * f * \hat{x}) \quad (2.13)$$

$$\hat{x} = x * \cos(\theta) + y * \sin(\theta) \quad (2.14)$$

$$\hat{y} = y * \cos(\theta) - x * \sin(\theta) \quad (2.15)$$

avec  $\theta$  désigne l'orientation de la sinusoidale, sa fréquence  $f$  et  $\sigma_x, \sigma_y$  l'ecart-type de la gaussienne selon l'axe des abscisses et des ordonnées respectivement. Pour appliquer les filtres de Gabor sur une image, les quatre paramètres doivent être spécifiés. Après l'application de ce filtre sur l'image, on calcule les paramètres de texture. Parmi ces paramètres, on peut citer la moyenne, la variance, l'entropie et l'énergie [88].

### 3.2. Ondelettes

Les ondelettes ont été introduites dans les travaux de Daubechies [89], Mallat [90] et Meyer [91]. La première définition de la transformée en ondelettes a été établie par Grossmann et Morlet [92] en temps et échelle continus. La théorie des ondelettes repose sur l'idée que le signal peut être caractérisé par différentes échelles et différentes résolutions. Une étude plus détaillée sur l'application en ondelettes. Se trouvera dans les références [93, 94, 95, 96].

#### 3.2.1. La transformée en ondelettes continue

La transformation en ondelettes continue est définie comme suit :

$$\tilde{f}(a, b) = \int_{-\infty}^{+\infty} f(t)\psi_{a,b}^*(t)dt \quad (2.16)$$

dans cette équation,  $\psi^*(t)$  est l'ensemble des fonctions de base appelées ondelettes. La famille d'ondelettes est générée à partir de la dilatation et de la translation d'une ondelette mère  $\psi(t)$ . Les ondelettes sont donc définies par :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (2.17)$$

L'équation ci-dessus devient :

$$\tilde{f}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad (2.18)$$

Où  $\psi$  désigne la fonction ondelette mère dont les dilatées et translatées sont les bases d'un espace d'analyse en ondelettes sur lequel est projetée la fonction  $f(t)$  [97].

La transformée en ondelettes est réversible, on peut donc passer de l'analyse d'une fonction à sa reconstruction par :

$$f(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{f}(a, b) \psi_{a,b}(t) \frac{da}{a^2} db \quad (2.19)$$

$\psi$  doit posséder certaines caractéristiques pour que la transformée en ondelettes soit possible. Toute fonction qui vérifie la condition d'admissibilité ci-dessous est une ondelette potentielle [97] :

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty \quad (2.20)$$

C'est-à-dire que l'ondelette est à énergie finie. De plus, on a  $|\hat{\psi}(0)|^2 = 0$ , autrement dit la fonction  $\psi$  est un filtre passe bande. Cela implique que

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0 \quad (2.21)$$

donc  $\psi$  est de moyenne nulle, elle doit osciller,  $\psi$  est une onde.

### 3.2.2. La transformée en ondelettes discrète

Les coefficients de la transformée en ondelettes discrète représentent la projection du signal par les fonctions de base générées par la translation et la dilatation de l'ondelette mère. La sélection de cette dernière détermine la représentation du signal. Les différentes ondelettes mères présentent des performances différentes dépendamment des caractéristiques du signal [98, 99].

— Bases d'ondelettes orthogonales

Pour numériser un signal continu  $x(t)$  les valeurs de dilatation et de translation  $a$  et  $b$  de l'ondelette même doivent être discrètes. Elles sont représentées comme suit,

$$a = a_0^j \text{ et } b = kb_0 a_0^j \quad (2.22)$$

Avec,  $a_0 > 1$  et  $b_0 > 0$  et  $j=0,1,2,\dots$ , et  $k$  constante.

En remplaçant l'équation (2.22) dans l'équation (2.17), l'ondelette mère dilatée et translatée s'écrit,

$$\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}} \psi\left(\frac{t - kb_0 a_0^j}{a_0^j}\right) \quad (2.23)$$

Ainsi, la transformée discrète en ondelettes [100] est écrite comme suit :

$$A_{j,k}(t) = \frac{1}{\sqrt{a_0^j}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{t - kb_0 a_0^j}{a_0^j}\right) dt \quad (2.24)$$

— Algorithme rapide de décomposition multi-résolution

L'algorithme rapide de calcul de la transformée en ondelettes a été introduit par Mallat en 1989 [101]. Cet algorithme est connu sous le nom de codeur à sous bande à deux canaux et il utilise des filtres miroirs quadratiques. La décomposition du signal  $x(n)$  produit deux types de coefficients : coefficients d'approximation et coefficients de détail. Ces vecteurs sont obtenus par la convolution du signal d'entrée avec les filtres passe-bas et passe-haut respectivement suivis par une décimation ( voir figure 2.4) ci-dessous :

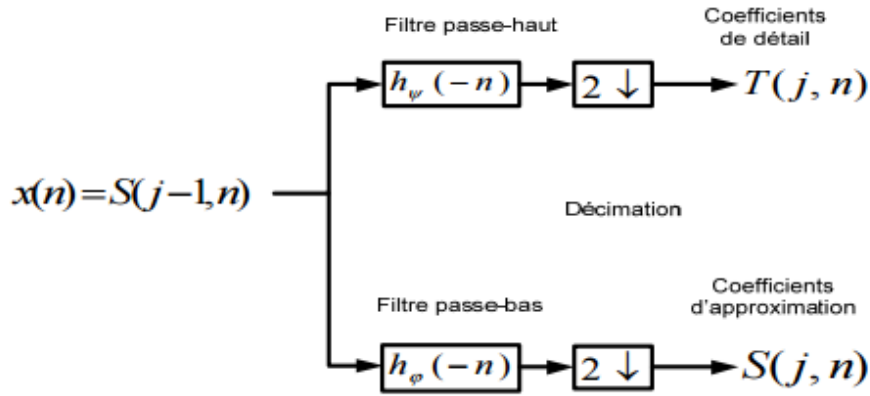


FIGURE 2.4 – Décomposition du signal  $x(n)$

— Coefficients d'approximation :

Les coefficients d'approximation  $S(j,k)$  sont déterminés à partir de la fonction d'échelle  $\varphi_{j,k}(t)$  et du signal  $x(t)$  suivant la formule de l'équation (2.25).

$$S_{j,k}(t) = \int_{-\infty}^{+\infty} x(t) \varphi_{j,k}(t) dt \quad (2.25)$$

avec :

$$\varphi_{j+1,k}(t) = \frac{1}{\sqrt{2}} \sum_{u=-\infty}^{+\infty} c_u \varphi_{j,2k+u}(t) \quad (2.26)$$

En remplaçant  $\varphi_{j,k}(t)$  par l'équation (2.26), on aura :

$$S(j, k) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{2}} \sum_{u=-\infty}^{+\infty} c_u \varphi_{j-1,2k+u}(t) dt \quad (2.27)$$

$$S(j, k) = \frac{1}{\sqrt{2}} \sum_{u=-\infty}^{+\infty} c_u \int_{-\infty}^{+\infty} x(t) \varphi_{j-1,2k+u}(t) dt \quad (2.28)$$

$$S(j, k) = \frac{1}{\sqrt{2}} \sum_{u=-\infty}^{+\infty} c_u S(j-1, 2k+u) \quad (2.29)$$

Si on pose,  $2k+u = v$ , alors :

$$S(j, k) = \frac{1}{\sqrt{2}} \sum_{v=-\infty}^{+\infty} c_{v-2k} S(j-1, v) \quad (2.30)$$

— Coefficients de détail :

Aussi, les coefficients de détail  $T(j,k)$  sont déterminés à partir de la fonction d'ondelettes  $\psi_{j,k}(t)$  et du signal  $x(t)$  suivant la formule suivante, [98]

$$T(j, k) = \int_{-\infty}^{+\infty} x(t) \psi_{j,k}(t) dt \quad (2.31)$$

$$\psi_{j+1,k}(t) = \frac{1}{\sqrt{2}} \sum_{u=-\infty}^{+\infty} b_u \psi_{j,2k+u}(t) \quad (2.32)$$

En remplaçant  $\psi_{j,k}(t)$  par l'équation (2.32), et en suivant les mêmes étapes de l'équation (2.27) à l'équation (2.30), on aura :

$$T(j, k) = \frac{1}{\sqrt{2}} \sum_{v=-\infty}^{+\infty} b_{v-2k} S(j-1, v) \quad (2.33)$$

Ici, les coefficients de détail de l'itération  $j$  sont dépendants des coefficients de détail de l'itération précédente  $j-1$ . À partir des deux équations (2.30) et (2.33), on voit bien que si les coefficients de l'approximation  $S(j,k)$  sont connus à l'instant de décomposition  $j=0$ , alors l'itération des équations ci-dessus de  $S(j,k)$  et  $T(j,k)$  génèreront les coefficients d'ondelette d'approximation et de détail à tous les instants de décomposition supérieures à  $j=0$ . Le calcul itératif de  $S(j,k)$  et de  $T(j,k)$  représente respectivement le filtrage passe-haut et le filtrage passe-bas du signal d'entrée.

La figure (2.5) représente la décomposition en ondelettes du signal  $x(n)$  avec une itération égale à trois. On dénote le vecteur d'entrée  $X_0$ , les vecteurs d'approximation  $S_j$ , et les vecteurs de détails  $T_j$

— Signal bidimensionnel :

Pour réaliser la transformation en ondelettes sur une image, on pourrait appliquer le même algorithme de décomposition multirésolution de Mallat, utilisé pour un signal monodimensionnel. Le calcul de la transformée discrète en ondelettes d'une image est la transformée discrète à une dimension des lignes suivie par la transformée discrète à une dimension des colonnes. La figure (2.6) visualise la décomposition d'une image en quatre images, trois de détails et une d'approximation.

La séquence d'entrée de l'image est décrite sous forme matricielle  $S_0$ . Nous avons besoin de trois bancs de filtres pour effectuer l'analyse de l'image à une itération. Le premier banc de filtres est utilisé pour la transformation niveau ligne et les deux autres bancs de filtres sont utilisés pour la transformation niveau colonne. Le filtre passe-bas  $h_\varphi(n)$  et le filtre passe-haut  $h_\psi(n)$  sont convolués avec la matrice d'entrée niveau ligne puis chaque sortie des deux



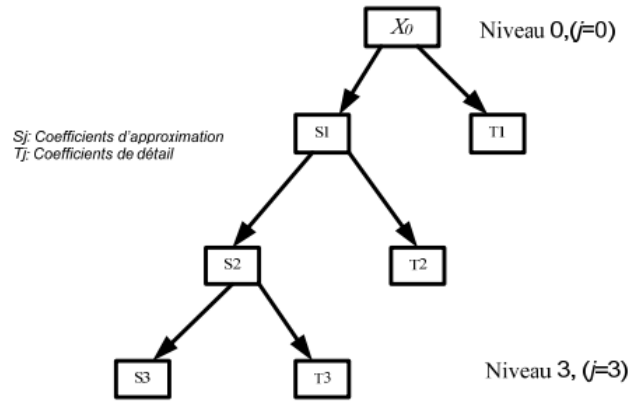


FIGURE 2.5 – Schéma bloc de la décomposition par ondelettes en trois itérations

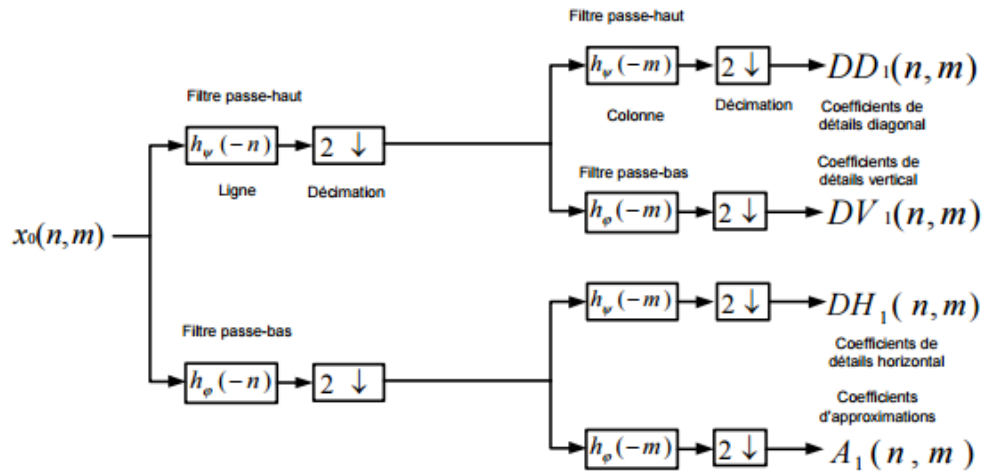


FIGURE 2.6 – Processus d'analyse multi-résolution à deux dimensions pour une itération

filtres est suivie d'une décimation d'un facteur de 2. Les deux séquences résultantes niveau colonnes vont aussi convoluées avec les deux autres bancs de filtres identiques au premier banc de filtres pour donner trois séquences de détails diagonal (DD1), vertical (DV1) et horizontal (DH1), et une séquence d'approximation (A1). C'est la transformation séparable niveau ligne colonne. La convolution de la matrice niveau ligne avec  $h_\varphi(-n)$  engendre la séquence suivante,

$$[x_0(n, m) * h_\varphi(-n)] \downarrow 2 \quad (2.34)$$

La convolution de l'équation (2.34) niveau colonne avec  $h_\varphi(-m)$  engendre la séquence d'approximation

$$A1 = [[[x_0(n, m) * h_\varphi(-n)] \downarrow 2] * h_\varphi(-m)] \downarrow 2 \quad (2.35)$$

La convolution de l'équation (2.34) niveau colonne avec  $h_\psi(-m)$  engendre la séquence de détail horizontal

$$DH1 = [[[x_0(n, m) * h_\varphi(-n)] \downarrow 2] * h_\psi(-m)] \downarrow 2 \quad (2.36)$$

La convolution de la matrice niveau ligne avec  $h_\psi(-n)$  engendre la séquence suivante,

$$[x_0(n, m) * h_\psi(-n)] \downarrow 2 \quad (2.37)$$

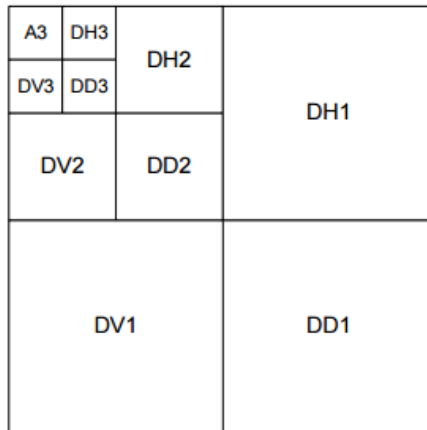
La convolution de l'équation (2.37) niveau colonne avec  $h_\varphi(-m)$  engendre la séquence de détail vertical

$$DV1 = [[[x_0(n, m) * h_\psi(-n)] \downarrow 2] * h_\varphi(-m)] \downarrow 2 \quad (2.38)$$

La convolution de l'équation (2.37) niveau colonne avec  $h_\psi(-m)$  engendre la séquence de détail diagonal

$$DD1 = [[[x_0(n, m) * h_\psi(-n)] \downarrow 2] * h_\psi(-m)] \downarrow 2 \quad (2.39)$$

Pour une deuxième, troisième ou quatrième itération c'est uniquement les signaux d'approximation à deux dimensions qui sont décomposés en quatre autres signaux .




---

FIGURE 2.7 – Décomposition d'une image en trois itérations

La figure (2.7) présente un schéma bloc de la décomposition d'une image en trois itérations. La première itération de l'image originale produit l'image DD1, DV1, DH1 et

A1. La deuxième itération permet la décomposition de l'image A1 en DD2, DV2, DH2 et A2. Enfin la troisième itération permet la décomposition de l'image A2 en DD3, DV3, DH3 et A3.

— La transformation inverse

La transformation inverse ou la phase de synthèse pour une itération nécessite trois bancs de filtres et trois sommateurs, comme présenté dans la figure (2.8). Cette transformation est aussi séparable niveau ligne colonne. On peut voir dans la figure (2.8) que les quatre séquences de sortie de la phase d'analyse (la séquence d'approximation et les trois séquences de détails) sont interpolées d'un facteur de 2 puis convoluées aux filtres correspondant en parallèle. Les sorties des filtres sont sommées deux à deux, la séquence d'approximation avec la séquence de détail horizontal et la séquence de détail vertical avec la séquence de détail diagonal. À ce niveau, les deux séquences résultantes vont aussi être interpolées et convoluées aux filtres correspondants en parallèle et finalement sommées pour donner une image reconstruite  $Y(n,m)$  [100].

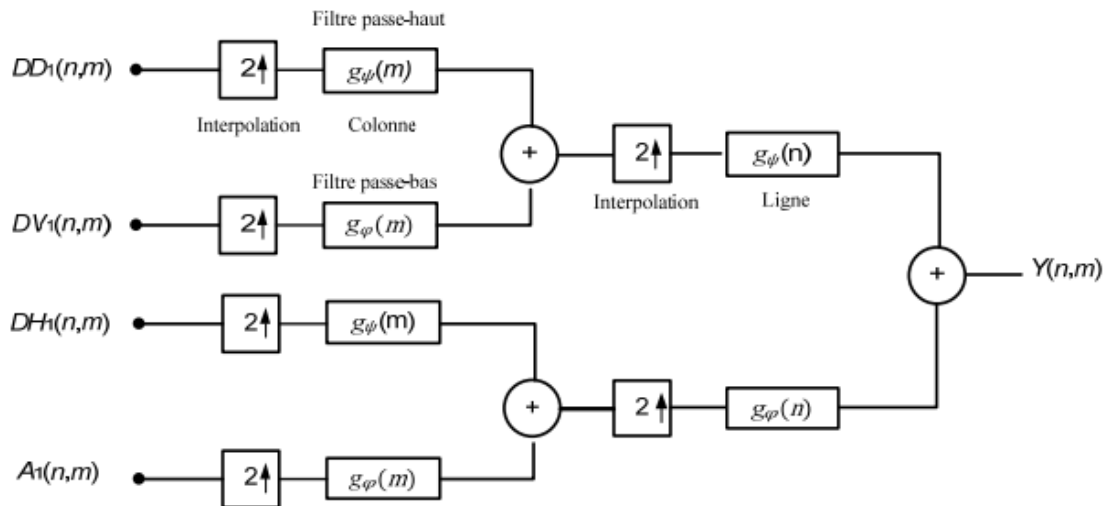


FIGURE 2.8 – Processus de reconstruction à deux dimensions pour une itération

### 3.2.3. Les familles de la transformée en ondelettes

Nombreux chercheurs ont construit des familles d'ondelettes possédant certaines propriétés remarquables. Parmi ces familles d'ondelettes on peut citer les ondelettes de Haar qui sont les plus simples, mais elles ne sont pas bien localisées. Ingrid Daubechies a construit des ondelettes à support compact qui permettent d'utiliser des filtres de taille finie. Une autre famille d'ondelettes est la famille des ondelettes splines dont la réponse fréquentielle est bien localisée. Les différentes familles d'ondelettes sont utilisées selon leurs propriétés en fonction du problème à résoudre. Dans la figure (2.9) nous présentons quelques familles d'ondelettes.

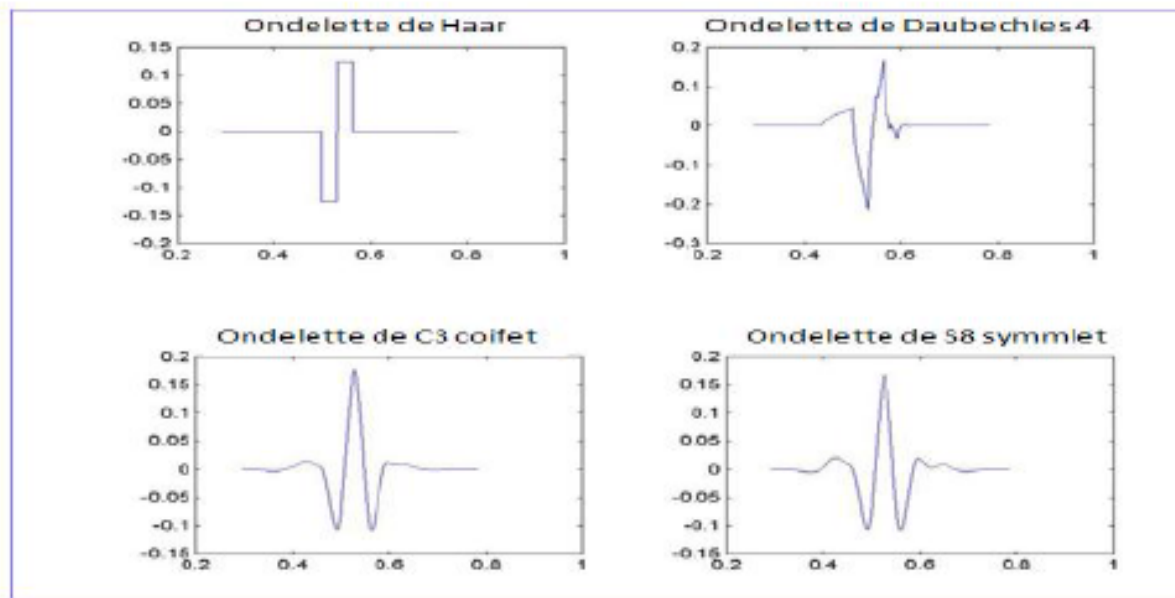


FIGURE 2.9 – Quelques familles d’ondelettes

### 3.3. La transformée en Contourlet

La transformée en Contourlet a été introduite par Minh Do [102] et Martin Vetterli [103], est un nouveau schéma de décomposition d’images. Cette décomposition est construite en combinant successivement deux étages de décomposition : une décomposition multi échelle suivie d’une décomposition directionnelle. Le premier étage utilise une pyramide laplacienne pour transformer l’image en une suite de niveaux LP passe-bande et un niveau passe-bas (approximation en basse fréquence de l’image). Le deuxième étage applique des filtres bi-dimensionnels pour décomposer chaque niveau LP passe-bande en un nombre de bandes directionnelles, capturant ainsi des informations directionnelles ou orientées. Finalement, l’image se trouve représentée par un ensemble de sous bandes multi échelles et orientées.

Comparée à la transformée en ondelette discrète, la transformée en contourlet avec sa propriété de sélectivité directionnelle conduit à des améliorations et à de nouveaux potentiels pour les applications de traitement d’image. En effet, les contours fins sont mieux représentés puisque des expérimentations ont déjà clairement montré que les contours lisses sont représentés de manière efficace par quelques coefficients contourlet localisés dans la bande à orientation appropriée [3].

### 3.4. La transformée en contourlet non sous échantillonnée

La transformée en Contourlets non sous échantillonnée, NSCT (NonsubSampled Contourlet Transform) [104], est une version invariante par translation de la transformée en Contourlets. Pour assurer l’invariance par translation, la NSCT est implantée à l’aide d’une structure pyramidale non sous-échantillonnée et des bancs de filtres directionnels non sous-échantillonnés. La figure (2.10) illustre le principe de la NSCT. Il s’agit de combiner successivement deux étages de décomposition invariants par translation [104] : (1) une structure

pyramidale non sous-échantillonnée, NSP (NonSubsampled Pyramid), qui assure la propriété multiéchelle et (2) une structure de bancs de filtres directionnels non sous-échantillonnés, NSDFB (NonSubsampled Directional Filter Bancs), permettant une décomposition suivant différentes orientations.

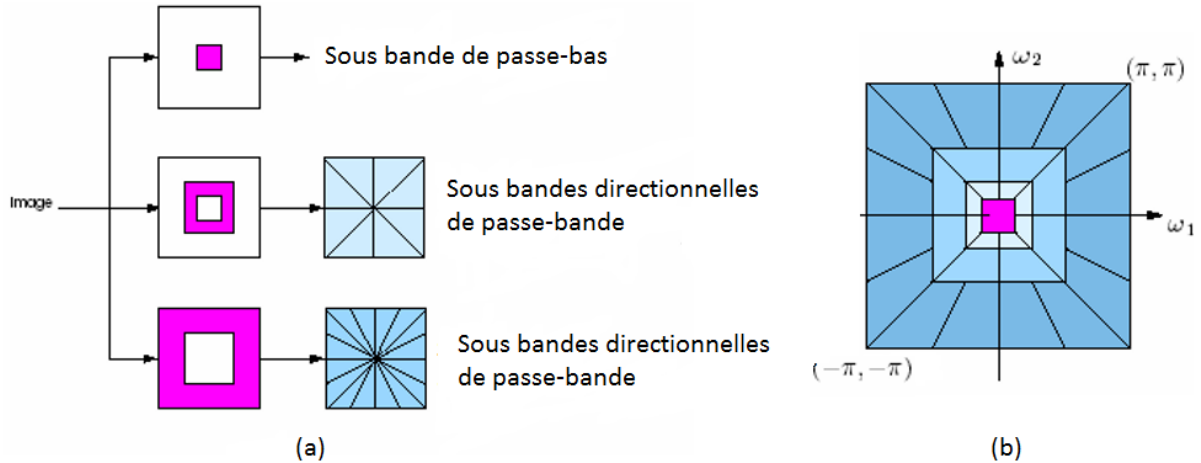


FIGURE 2.10 – Transformation en contourlets non sous-échantillonnée : a) Schéma bloc de l'implantation de la NSCT à l'aide d'un banc de filtre pyramidal directionnel non sous-échantillonné, b) Banc de filtre directionnel [3].

### 3.4.1. Pyramide non sous échantillonnée

Différente de la transformée en contourlettes, l'étape de décomposition multiéchelle de la NSCT est réalisée par une structure de filtrage invariante au décalage, qui réalise une décomposition de sous-bande semblable à celle de la pyramide laplacienne, en employant des bancs de filtres bidimensionnels non sous-échantillonnés à deux canaux. En raison du non sous-échantillonnage dans la pyramide de décomposition, la sous-bande passe-bas n'a aucun crénelage de fréquence, le crénelage peut arriver lorsqu'on redimensionne une image vers une résolution différente de la résolution de numérisation. Il s'ensuit un effet de pixellisation, ou aliasing, sur les bordures entre objets, plus ou moins marqué. De plus, même la largeur de bande du filtre passe-bas est plus grande que celle utilisée par les contourlettes. Par conséquent, la NSCT a de meilleures caractéristiques fréquentielles que la transformée en contourlettes.

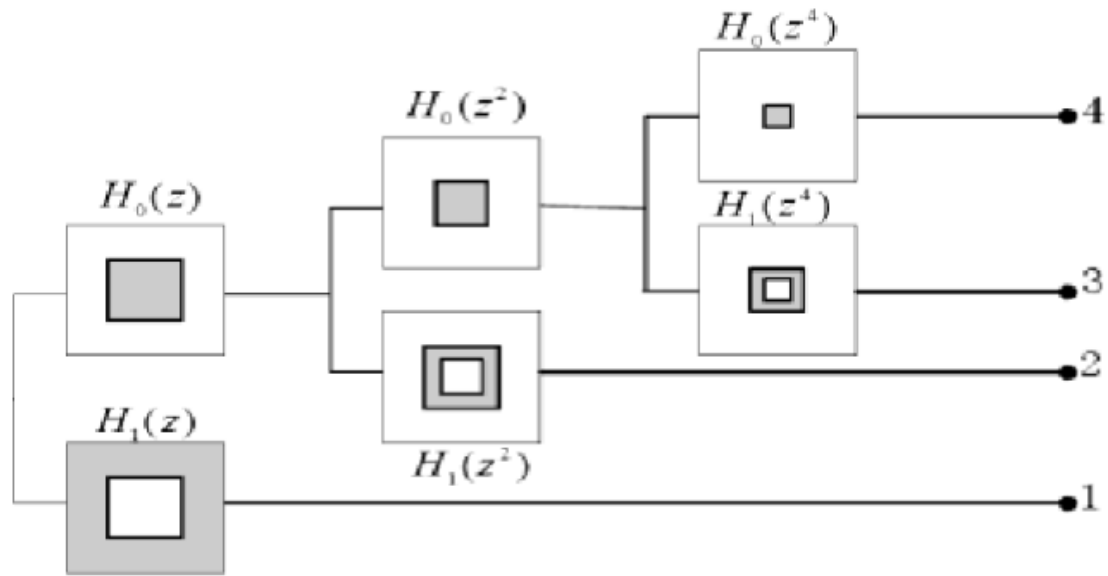


FIGURE 2.11 – Pyramide de décomposition non sous échantillonnée à 3 niveaux.

Le  $H_0(Z)$  est un filtre passe-bas et  $H_1(Z)=1-H_0(Z)$  . La condition de reconstruction parfaite est donnée par :  $H_0(Z)*G_0(Z)+H_1(Z)*G_1(Z)=1$ . Ou  $G_0(Z)$  et  $G_1(Z)$  sont des filtres passe-bas et passe-haut. La figure (2.11) donne la décomposition de la pyramide non-sous-échantillonnée à 3 niveaux. La pyramide transforme l'image en une suite de niveaux passe-haut et un niveau passe-bas. Les pyramides non-sous-échantillonnées sont construites en itérant des bancs de filtres non-sous-échantillonnées. Les filtres des niveaux ultérieurs sont obtenus en sur-échantillonnant les filtres du premier niveau. Ceci donne la propriété de multi-échelle sans besoin de concevoir des filtres additionnels. La condition de reconstruction parfaite est ainsi plus facile à satisfaire qu'avec un échantillonnage critique (exemple : contourlettes, directionlettes...), et permet de concevoir des filtres meilleurs.

### 3.4.2. Bancs de filtres directionnels non-sous-échantillonnés

Après la décomposition en sous-bandes passe-haut et passe-bas par la pyramide non-sous-échantillonnée, les bancs de filtres directionnels (DFB) non sous échantillonnées décomposent la sous-bande passe-haut en plusieurs sous-bandes directionnelles, en appliquant de manière appropriée des filtres bidimensionnels non sous échantillonnées à deux-canaux de type quin-conce pour décomposer chaque niveau passe-haut en un nombre de bandes directionnelles, capturant ainsi des informations directionnelles ou orientées.

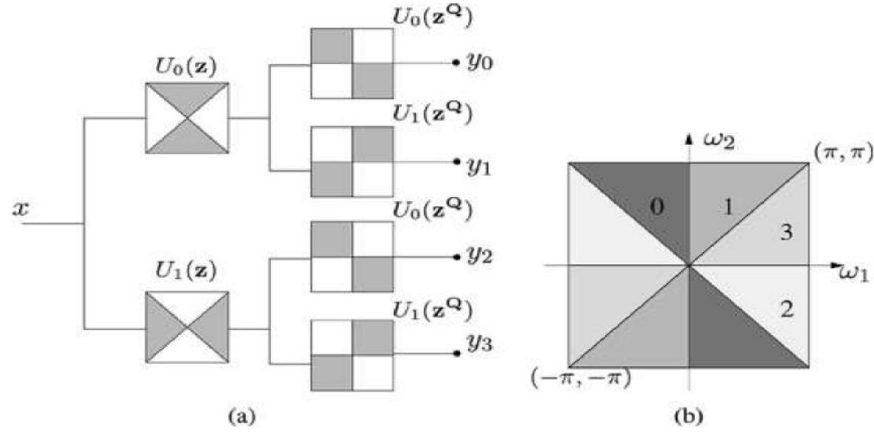


FIGURE 2.12 – Bancs de filtres directionnels non-sous-échantillonnés construits avec des bancs de filtres en éventail à deux-canaux.

Les bancs de filtres directionnels non-sous-échantillonnés sont une version invariante par translation (shift-invariant) des bancs de filtres directionnels à échantillonnage critique de la transformée en contourlettes. Ils sont construits en éliminant le sous-échantillonnage et le sur-échantillonnage dans chaque banc de filtres à deux-canaux de la structure arborescente de la DFB et en suréchantillonnant les filtres respectifs. Cela aura comme résultat un arbre composé de bancs de filtres non-sous-échantillonnés à deux-canaux [105]. La décomposition directionnelle est obtenue en réitérant les bancs de filtres directionnels non-sous-échantillonnés. Tous les niveaux ultérieurs sont par la suite sur-échantillonnés par une matrice quinconce afin de compenser le non sous échantillonnage. Les réponses fréquentielles de deux filtres sur-échantillonnés est donné par la figure (2.13), et la cascade d'analyse est montrée dans la figure (2.12.a), où des bancs de filtres directionnels non-sous-échantillonnés à quatre-canaux sont construits avec des bancs de filtres en éventail à deux-canaux. Nous obtenons par la suite une division fréquentielle à quatre-directions comme montré dans la figure (2.12.b). La NSCT est ainsi construite en combinant la pyramide non-souséchantillonnées et les bancs de filtres directionnels non-sous-échantillonnés, tel que montré dans la figure (2.14). La pyramide non-sous-échantillonnée fournit une décomposition multi-échelle en divisant l'entrée en sous-bande passe-bas et passe-haut, et les bancs de filtres directionnels non-sous-échantillonnés fournissent une décomposition directionnel en divisant la sous-bande passe-haut en plusieurs sous-bandes directionnels. Ce schéma peut être répété sur la sous-bande passebas, à la sortie de la pyramide non-sous-échantillonnée.

Le noyau de la NSCT est les bancs de filtres non-sous-échantillonnés inséparables à deux-canaux. Il est plus facile et plus flexible de concevoir les bancs de filtres nécessaires qui mènent à une transformée en contourlettes non-souséchantillonnée, avec une meilleure sélectivité et régularité fréquentiel, quand cela est comparé à la transformée en contourlettes. Les éléments de structure de la NSCT sont d'autant réguliers et symétriques.

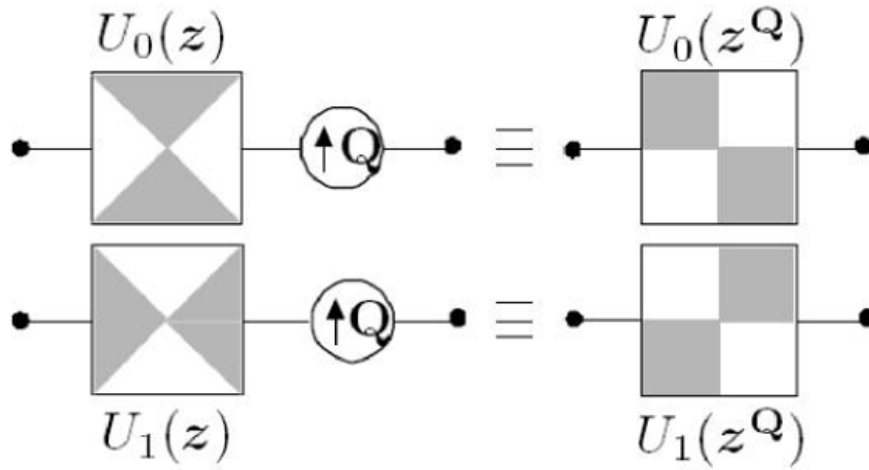


FIGURE 2.13 – Filtres sur-échantillonnés par une matrice quinconce  $Q$

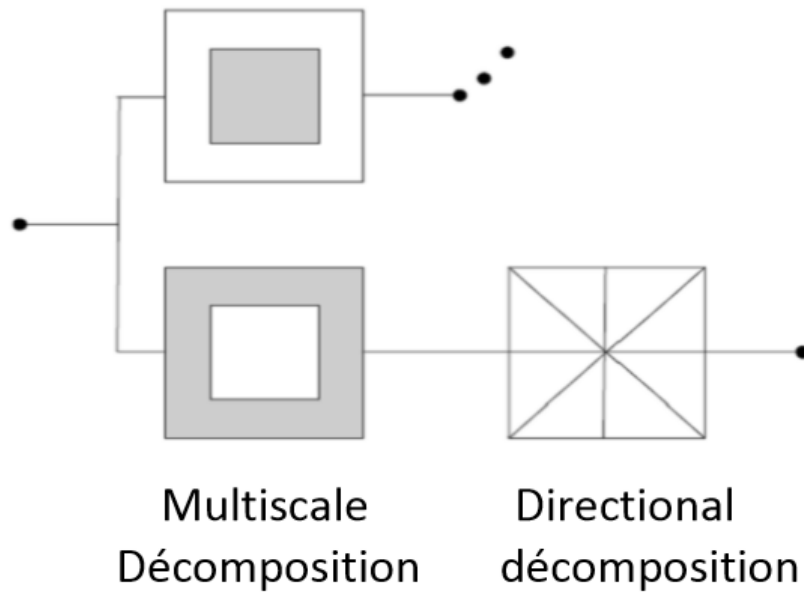


FIGURE 2.14 – Le bloc de diagramme de la NSCT



### 3.5. La morphologie mathématique

La morphologie mathématique est un outil essentiellement utilisé en traitement d'images, pour étudier les objets en fonction de leur (forme, taille, voisinage, texture, niveaux de gris...). Elle touche les différentes étapes du traitement d'images (analyse de texture, filtrage, segmentation). On la utilise dans de nombreux domaines du traitement d'images comme (biologie, imagerie médicale, satellitaire, robotique).

Le principe de cette théorie, consiste à comparer une forme à analyser avec une forme appelée élément structurant. Cette opération permet en chaque point d'effectuer une comparaison à travers des relations telles que l'union, l'intersection, la complémentation et l'inclusion.

— La dilatation

Soit  $X$  un sous-ensemble de  $E$ . La dilatation morphologique de  $X$  par un élément structurant  $B$ , noté  $D^B(X)$ , est définie par l'ensemble des points  $x$  de l'espace pour lesquelles  $B_x$  présente une intersection non vide avec  $X$  :

$$D^B(X) = \{x \in E : B_x \cap X \neq \emptyset\} \quad (2.40)$$

— L'érosion

L'érosion d'un ensemble  $X$  par un élément structurant  $B$ , noté  $E^B(X)$ , est l'ensemble des points  $x$  de l'espace pour lesquels  $B_x$  est contenu dans  $X$  :

$$E^B(X) = \{x \in E : B_x \subseteq X\} \quad (2.41)$$

— Fermeture

La fermeture morphologique d'un ensemble  $X$ , noté  $X \bullet B$ , est correspond à la dilatation de  $X$  par  $B$  suivie de l'érosion par  $B$  :

$$X \bullet B = E^B(D^B(X)) \quad (2.42)$$

— L'ouverture

L'ouverture morphologique d'un ensemble  $X$ , noté  $X \circ B$ , est correspond à l'érosion de  $X$  par  $B$  suivie de la dilatation par  $B$  :

$$X \circ B = D^B(E^B(X)) \quad (2.43)$$

— Gradients

Les gradients morphologiques sont les résidus de la dilatation et de l'érosion morphologique :

$$Grad_B(X) = D^B(X) - E^B(X) \quad (2.44)$$

— Chapeau haut de forme (Top-Hats)

F. Meyer a ajouté une transformation appelée "chapeau haut de forme" pour détecter ce qu'a éliminé une ouverture ou ce qu'a rajouté une fermeture par rapport à l'image originale.

c-à-d pour extraire les pics clairs de luminance on définit le chapeau haute forme blanc ou (WTH) (white top hat, en anglais). Elle est définie comme la différence algébrique entre l'identité ( $f$ ) et son ouverture  $O^B(f)$  tel que :

$$WTH^B(f) = f - O^B(f) \quad (2.45)$$

Par symétrie, pour extraire les vallées sombres de l'image on définit le chapeau haute forme noir ou (BTH) (black top hat, en anglais) est défini comme la différence algébrique entre la fermeture  $F^B(f)$  et l'identité ( $f$ ) tel que :

$$BTH^B(f) = F^B(f) - f \quad (2.46)$$

## 4. Segmentation du muscle pectoral

Sur une mammographie celle présentée sur la figure (1.2)(voir la page 5), nous pouvons observer le muscle pectoral qui représente une vue de la région de densité prédominante de muscle pectoral apparaît approximativement au même densité que les tissus denses d'intérêt dans l'image mammographie ce qui peut affecter les résultats des méthodes de traitement d'image [106]. Dans la littérature, de nombreuses méthodes de segmentation des muscles pectoraux existent. Dans ce qui suit, nous citons quelques travaux :

Karssemeijer et al [107, 108] utilisent la transformée de Hough et un ensemble de valeurs de seuil appliqués afin de détecter le muscle pectoral. Ferrari et al. [109] présentent une technique basée sur la banque de filtre à ondelettes de Gabor. Mirzaalian et al [110] présentent une approche pour la segmentation des muscles pectoraux basé sur l'algorithme de diffusion non linéaire. Mario Mustra et al [111] proposent une méthode qui utilise la décomposition en ondelettes pour trouver la frontière du muscle pectoral. A. Boucher et al [112] développent une méthodologie basée sur les contours actifs, où les différentes énergies sont proposées. KS Camilus [113] proposent la technique de graph-cut pour la segmentation d'image qui permet d'identifier le bord du muscle pectoral. Kwok SM et al [114] utilisent un algorithme adaptatif; pour la connaissance de la position et la forme du muscle pectoral. Le bord pectoral est d'abord estimée par une ligne droite qui est validé pour l'exactitude de l'emplacement et l'orientation. Raba et al [115] combinent un algorithme du croissante de région et l'approche d'histogramme adaptative pour atteindre la segmentation du muscle pectoral. Ma. F, Bajger et al [116] présentent deux méthodes fondées sur les graph-cut qui repose sur des informations d'image globale pour la segmentation du muscle pectoral. La première méthode utilise l'algorithme du Minimum Spanning Tree (MST) qui est bien connu pour sa vitesse, et la seconde méthode utilise l'algorithme Adaptative Pyramide (AP) pour la détection du limite pectoral. Ces deux méthodes utilisent la technique du contour active en commun pour améliorer les détails de la frontière du muscle pectoral. J. Chakraborty et al [4] proposent une nouvelle approche pour la détection du muscle pectoral en utilisant la moyenne du gradient.

## Conclusion

Dans ce chapitre, nous avons présenté différents travaux citer dans la littérature concernant les différentes étapes de Diagnostic Assisté par Ordinateur (DAO) à savoir la phase de prétraitement des images mammographiques, la segmentation des microcalcifications ou les lésions mammaires.

Le chapitre suivant se focalisera à une étude détaillée sur les descripteurs de la texture, les classificateurs et les techniques de l'évaluation de la performance de la classification. Le

choix d'un meilleur ensemble de descripteurs de texture promet une bonne classification des images mammographiques.

# Chapitre 3

## Classifications des images mammographiques

### Introduction

Pour toute application visant l'utilisation des textures pour la classification, la segmentation, la recherche d'images ou autres, elle nécessite une phase de la description à pour but d'extraire les caractéristiques ou les paramètres de textures qui décrivent les objets présents dans l'image. Il s'agit dans cette étape de transformer les images de la base en paramètres numériques ayant une signification physique. L'extraction des descripteurs de texture permet de projeter les images texturées dans un espace ayant une dimension importante.

Ainsi que les méthodes de description d'images sont variables et dépendent du type d'image à analyser (image binaire, image en niveau de gris, image couleur) et aussi l'extraction des caractéristiques dans le cas des microcalcifications est une étape essentielle [117, 118, 119]. Parmi ces caractéristiques, on peut en trouver qui portent sur la texture comme l'analyse des régions avoisinantes, la dépendance spatiale des niveaux de gris, les statistiques de longueurs sur les niveaux de gris ou encore la différence de niveaux de gris [120, 121, 122].

Une autre classe de mesures repose sur l'analyse de texture [123]. Ainsi on peut citer les mesures dérivées de matrices de co-occurrence des niveaux de gris [124, 125, 126, 127].

La phase d'extraction est la phase clé dans la classification des images mammographiques. Un attribut est défini comme la connaissance utilisée pour caractériser l'information contenue dans les images [128]. Nous distinguons deux types de descripteurs [129] :

1-) Descripteur bas-niveau : il représente l'information visuelle de bas telle que la couleur, la texture, la forme ou l'information spatiale [128].

2-) Descripteur haut-niveau : décrit le contenu sémantique [129] de l'image et principalement représenté par des mots clés fournis par l'utilisateur, car l'extraction automatique est un problème non résolue actuellement. Dans ce travail, nous intéressons à la texture.

## 1. Les descripteurs de la couleur

La couleur est un descripteur robuste qui est très utilisé pour l'apprentissage ou la recherche d'images [130], la forme représentative de la couleur la plus utilisée est la technique de l'histogramme qui présente la distribution complète de la couleur dans une image, mais malgré sa robustesse, il pose des problèmes de perte de temps et de saturation de la mémoire vu ces problèmes les moments statistiques de l'histogramme sont utilisés pour décrire la distribution de la couleur [131] en calculant seulement les caractéristiques dominantes [132, 128].

## 2. Les descripteurs de la forme

La forme est aussi un descripteur très important dans la recherche des images. La forme désigne l'aspect général d'un objet, son contour, il existe deux famille de représentation de la forme : une approche basée sur les contours représente les frontières externe de l'objet et une approche basée sur les régions, la plupart des travaux s'intéresse à la seconde approche [133]. Les algorithmes les plus efficaces pour les deux catégories utilisent les descripteurs de Fourier (plusieurs variantes existent) et les moments invariants (les moments de Hu par exemple) [134, 135].

## 3. Les descripteurs de la texture

La texture est le second descripteur visuel utilisé pour l'images [136], c'est la première caractéristique perçue par l'oeil humaine [137]. La texture est définie par une répétition d'un motif créant une image visuellement homogène, il contient une importante des informations sur l'organisation des primitives. De nombreuses approches de méthodes ont été proposées pour la caractérisation de la texture. Parmi les quelles nous citons : les méthodes statistiques et les méthodes fréquentielles.

### 3.1. Analyse des textures par les méthodes statistiques

Ce sont des méthodes basées sur la distribution de niveaux de gris [138]. Elles étudient les relations entre un pixel et son voisin. Elles sont utilisées pour caractériser les structures fines. Parmi ces méthodes, on peut citer la matrice de co-occurrence, les techniques de tamura et la méthode de différence de niveaux de gris [139].

#### 3.1.1. Technique de Tamura

L'approche de Tamura [140] est décrit les textures possibles [137] selon des concepts qui correspondent à la perception visuelle de l'humain. Les auteurs proposent six paramètres visuels de texture : la grossièreté, le contraste, la direction, la détection de ligne (linéarité), la régularité et la rugosité [141]. Ces paramètres sont calculés pour construire un vecteur de texture. Les trois premiers paramètres (la grossièreté, la direction, le contraste) ont une forte liaison avec la perception humaine [138].

— Grossièreté

est une mesure de la granularité de l'image [142]. Il se décompose en quatre étapes :

1. Pour chaque valeur  $x$  du pixel positionné en  $(i, j)$  :

$$A_K(x, y) = \frac{1}{2^{2k}} \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} x(i, j) \quad (3.1)$$

2. Ensuite, calculer respectivement les différences horizontales et verticales :

$$E_K^H(x, y) = |A_K(x + 2^k, y) - A_K(x - 2^k, y)| \quad (3.2)$$

$$E_K^V(x, y) = |A_K(x, y + 2^k) - A_K(x, y - 2^k)| \quad (3.3)$$

3. Puis, pour chaque pixel, choisir la valeur  $k$  qui maximise les différences horizontales et verticales :

$$\hat{K}(x, y) = \operatorname{argmax}_k \{E_K^H(x, y), E_K^V(x, y)\} \quad (3.4)$$

4. Enfin, la valeur du paramètre de grossièreté est obtenue comme la moyenne des tailles des meilleurs granules trouvés sur l'ensemble de l'image :

$$f_{crs} = \frac{1}{N \times M} \sum_{x=1}^N \sum_{y=1}^M \hat{K}(x, y) \quad (3.5)$$

#### — Contraste

Le contraste mesure la variation des niveaux de gris dans l'image ([142, 143]). Il est mesuré en fonction de l'écart-type  $\sigma$  et du kurtosis  $\alpha_4$  :

$$f_{con} = \frac{\sigma}{(\alpha_4)^\gamma} \quad (3.6)$$

ou  $\gamma$  est une constante positive. Selon les expériences faites par Tamura et al.,  $\gamma = \frac{1}{4}$  donne les meilleurs résultats.

#### — Direction

Le paramètre de direction mesure l'orientation totale dans l'image, les différentes directions qui peuvent éventuellement exister dans l'image ([142, 143]). Le calcul de ce paramètre s'effectue en quatre étapes :

1. Calculer pour chaque pixel  $p(i, j)$  de l'image, le gradient horizontal et vertical. Le gradient calculé sur un voisinage de pixels, reflète la direction des variations des niveaux de gris sur ce voisinage. Cela permet aussi de détecter les contours existants dans l'image. Tamura a calculé le gradient horizontal  $\Delta h$  en faisant la différence entre les trois niveaux de gris du côté gauche du pixel  $p(i, j)$  et les trois niveaux de gris du côté droit :

$$\Delta_h(i, j) = \sum_{k \in \{-1, 0, 1\}} (p(i + 1, j + k) - p(i - 1, j + k)) \quad (3.7)$$

Le gradient vertical  $\Delta v$  est la différence entre les trois niveaux de gris du dessus et du bas du pixel  $p(i, j)$  :

$$\Delta_v(i, j) = \sum_{k \in \{-1, 0, 1\}} (p(i + k, j + 1) - p(i + k, j - 1)) \quad (3.8)$$

2. Calculer la longueur  $e$  et l'angle d'orientation  $a$  sur chaque pixel comme suit :

$$e(i, j) = \frac{|\Delta_h(i, j)| + |\Delta_v(i, j)|}{2} \quad (3.9)$$

$$a(i, j) = \tan^{-1}\left(\frac{|\Delta_v(i, j)|}{|\Delta_h(i, j)|}\right) + \frac{\pi}{2} \quad (3.10)$$

3. Construire l'histogramme  $H$  en sommant tous les points dont la longueur  $e$  est supérieure à un certain seuil  $\alpha$  et en tenant compte de la valeur de l'angle  $a$  du contour. Plus précisément,  $H(k)$  est égal au nombre de pixels tels que :

$$\frac{2k-1}{2n} < \frac{a}{\pi} \leq \frac{2k+1}{2n} \pmod{1} \quad (3.11)$$

et  $|e| > \alpha$  avec  $k \in \{-1, 0, 1\}$

4. Une fois l'histogramme construit, la variance des valeurs des pics est calculée. La direction de l'image est calculée en sommant ces variances.

— Régularité

La régularité est donnée par :

$$f_{reg} = 1 - \alpha(\sigma_{crs} + \sigma_{con} + \sigma_{dir} + \sigma_{lin}) \quad (3.12)$$

— Rugosité

Le paramètre de rugosité est simplement obtenu en sommant les descripteurs de grossièreté et de contraste :

$$f_{rgh} = f_{crs} + f_{con} \quad (3.13)$$

### 3.1.2. Matrices de co-occurrences

Les caractéristiques de texture Haralick [144] sont utilisés pour la classification des images. Ces caractéristiques capturer des informations sur les motifs qui émergent dans les modèles de texture. Les caractéristiques sont calculées par la construction d'une matrice de co-occurrence. Certains de ces caractéristiques : le contraste, la corrélation, ainsi que d'une variété d'entropie

$$P_{0,d}(i, j) = |\{(n, m), (k, l) : n = k, |m - l| = d\}| \quad (3.14)$$

$$P_{45,d}(i, j) = \left| \left\{ \begin{array}{l} (n, m), (k, l) : n - k = d, m - l = -d \\ or(k - n = -d, l - m = d) \end{array} \right\} \right| \quad (3.15)$$

$$P_{90,d}(i, j) = |\{(n, m), (k, l) : |k - n| = d, l = m\}| \quad (3.16)$$

$$P_{135,d}(i, j) = \left| \left\{ \begin{array}{l} (n, m), (k, l) : (k - n = d, l - m = d) \\ or(k - n = -d, l - m = -d) \end{array} \right\} \right| \quad (3.17)$$

Selon les équations (3.14) à (3.17), les paramètres nécessaires pour le calcul du matrice de co-occurrences sont les suivantes :

— Distance entre pixels ( $d$ ) :

Normalement, les paires de pixels sont voisins, mais la matrice pourraient également être analyser par la relation entre les pixels non-consécutifs. Ainsi, une distance entre pixels doit être précédemment défini.

— angle ( $\theta$ ) :

Comme pour le paramètre de distance, il est nécessaire de définir la direction de la paire de pixels (voisins). Les indications les plus courantes sont 0, 45, 90, 135 et ses équivalents symétriques.

— Energie

L'énergie mesure l'homogenité globale de l'image, en effet, l'énergie a une valeur d'autant plus faible ceci indique qu'il y'a de nombreuse zones homogènes.

$$E = \sum_{i=0}^n \sum_{j=0}^n P_{d,\theta}(i, j)^2 \quad (3.18)$$

— Entropie

L'entropie permet de caractériser le degré de granulation de l'image, plus l'entropie est élevé plus la texture est dite granuleuse ou grossière, l'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire.

$$S = - \sum_{i=0}^n \sum_{j=0}^n P_{d,\theta}(i, j) \log P_{d,\theta}(i, j) \quad (3.19)$$

— Homogénéité

L'homogenité mesure si la texture possède des régions homogènes ou périodicité dans le sens de translation, lorsque l'homogenité est élevé cela indique que la texture possède des régions homogènes.

$$H = \sum_{i=0}^n \sum_{j=0}^n \frac{1}{1 + (i - j)^2} P_{d,\theta}(i, j) \quad (3.20)$$

— Uniformité

Ce paramètre mesure l'absence des différences niveaux de gris dans l'image. Sa valeur est d'autant plus élevée que les différences de niveaux de gris sont faibles.

$$U = \sum_{i=0}^n P_{d,\theta}(i, j)^2 \quad (3.21)$$

— Contraste

Ce paramètre mesure les variations locales des niveaux de gris, sa valeur est d'autant plus élevé ce implique que la texture possède moins de région homogène.

$$C = \sum_{i=0}^n \sum_{j=0}^n (i - j)^2 P_{d,\theta}(i, j) \quad (3.22)$$

— Direction

$$D = \sum_{i=0}^n P_{d,\theta}(i, j) \quad (3.23)$$

— Corrélation



Ce parametre mesure l'uniformité de la distribution de niveau de gris.

$$Cor = \frac{\sum_{i=0}^n \sum_{j=0}^n ij P_{d,\theta}(i, j) - \mu_i \mu_j}{\sigma_i \sigma_j} \quad (3.24)$$

avec :

$$\begin{aligned} \mu_i &= \sum_i \sum_j (i \cdot P(i, j)) \\ \mu_j &= \sum_i \sum_j (j \cdot P(i, j)) \\ \sigma_i &= \sum_i \sum_j (i - \mu_i)^2 P(i, j) \\ \sigma_j &= \sum_i \sum_j (j - \mu_j)^2 P(i, j) \end{aligned}$$

— Corrélation de l'information 1 et de l'information 2

$$corrinf1 = \frac{S - HXY1}{\max\{HX, HY\}} \quad (3.25)$$

$$corrinf2 = \sqrt{(1 - \exp[-2(HXY2 - HXY)])} \quad (3.26)$$

avec :

$$\begin{aligned} HXY &= - \sum_i \sum_j p(i, j) \log(p(i, j)) \\ HXY1 &= - \sum_i \sum_j p(i, j) \log\{p_x(i), p_y(j)\} \\ HXY2 &= - \sum_i \sum_j p_x(i) p_y(j) \log(p_x(i), p_y(j)) \end{aligned}$$

## 3.2. Analyse des textures par les méthodes fréquentielles

La texture est définie comme un mélange de signaux de fréquences, d'amplitudes et de directions différentes. Parmi les approches de description de la texture les plus utilisées concerne les propriétés fréquentielles, on peut citer la transformé de fourrier, le filtre de gabor, la transformée en ondelettes, les pyramides orientables, ainsi que les contourelets [145], [85, 146, 147].

### 3.2.1. L'énergie des ondelettes

L'énergie de chaque sous-bande est calculée puis moyennée par le nombre de coefficients de cette sous-bande :

$$E_k = \frac{1}{N_k} \sum_{m=1}^{N_k} (\omega_{k,m}^j)^2 \quad (3.27)$$

Où  $\omega_{k,m}^j$  est le  $m^{ième}$  coefficient d'ondelette de la  $k^{ième}$  sous-bande du niveau j.

k : indice de la sous-bande fréquentielle.

$N_k$  : le Nombre de coefficients de la  $k^{ième}$  sous-bande.

### 3.2.2. Le descripteur GIST

Dans la vision par ordinateur, les descripteurs de GIST sont une représentation d'une image en basse dimension qui contient suffisamment d'informations pour identifier la scène dans une image. Le descripteurs GIST est construit par Torralba [148, 149, 150] permet d'avoir une représentation très faible de la dimension d'une image. Les auteurs ont essayé de capturer le descripteur de GIST de l'image en analysant les fréquences spatiales et les

orientations tel que présenté dans la figure (3.1).

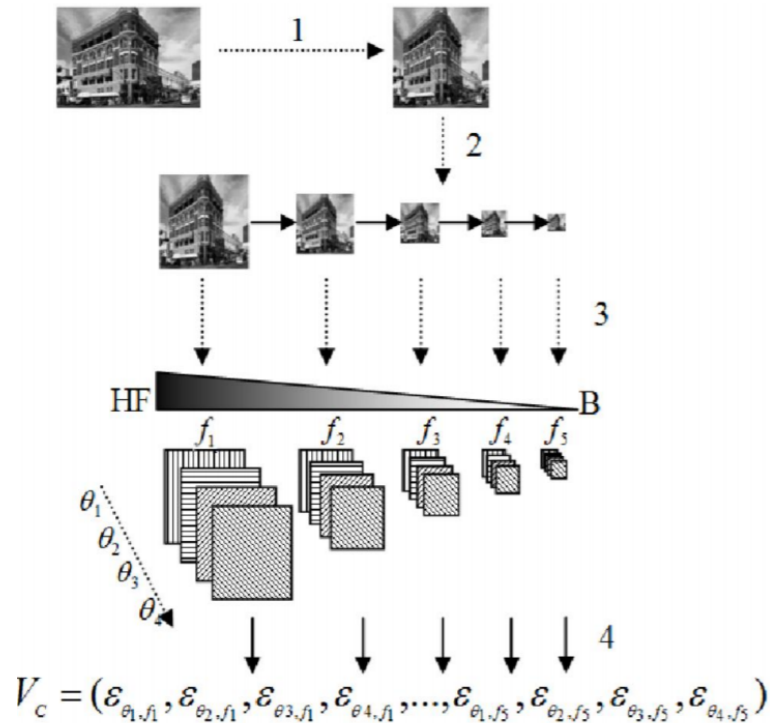


FIGURE 3.1 – Schéma de principe pour le calcul et l'extraction du descripteur de GIST

Comme précisé ci-dessus, la figure (3.1) présente un schéma résumant les différentes étapes pour le calcul et l'extraction de descripteurs de GIST. Après l'étape de prétraitement de l'image d'entrée, l'étape suivante consiste à modifier l'image en différentes échelles et orientations. Enfin, les vecteurs de caractéristiques de l'orientation et de la fréquence sont calculés pour chaque échelle. Ces vecteurs caractéristiques sont combinés pour former un descripteur global qui peut être réduit par une analyse en composantes principales (ACP). Le descripteur "brut" de GIST est construit par les étapes suivantes :

- On passe l'image dans un banc de filtres de Gabor [151] avec  $N_\sigma$  échelles et  $N_\theta$  orientations par échelle ; on obtient donc  $N = N_\sigma \times N_\theta$  images.
- Pour réduire la taille, chaque image est redimensionnée et subdivisée, on découpe chaque image en  $M \times M$  sous-images ( $M$  compris entre 2 et 16)
- On calcule l'énergie de chaque sous-image ; on obtient donc un vecteur de dimensions  $N \times M \times M$ .

Pour les images couleur, on procède ainsi pour chaque canal, le descripteur est donc de taille  $3 * N \times M \times M$ .

Nous utilisons les paramètres classiques :

$$N_\sigma = 4$$

$$N_\theta = 8$$

$$M = 4$$

La taille du vecteur est donc 512. Cette dimension peut être encore réduite par une analyse en composantes principales (ACP), qui donne également les poids appliqués aux différents

filtres [152].

## 4. L'apprentissage et la classification

La classification est considérée comme la dernière étape dans un système de diagnostic assisté par ordinateur, et a pour but de donner la réponse finale sur ce qui est détecté.

La prédiction est un processus à deux étapes : une étape d'apprentissage (entraînement) et une étape de classification (utilisation). Dans l'étape d'apprentissage, un classifieur (une fonction, un ensemble de règles, ...) est construit en analysant une base de données d'exemples d'entraînement avec leurs classes respectives. Un exemple  $X = (x_1, x_2, \dots, x_m)$  est représenté par un vecteur d'attributs de dimension  $m$ . Chaque exemple est supposé appartenir à une classe prédéfinie représentée dans un attribut particulier de la base de donnée appelé attribut de classe. Puisque la classe de chaque exemple est donnée, cette étape est aussi connue par l'apprentissage supervisé. Dans l'étape de classification, le modèle construit dans la première étape est utilisé pour classer les nouvelles données. Mais avant de passer à l'utilisation, le modèle doit être testé pour s'assurer de sa capacité de généralisation sur les données non utilisées dans la phase d'entraînement. On trouve plusieurs techniques de prédiction.

Une première famille de classifieurs comprend les méthodes de type plus proches voisins Cover et Hart [153]. L'idée de cette méthode est classifier un individu est de regarder les individus dont on connaît la classe et qui sont assez semblables. En mammographie, Veldkamp et Karssemeijer [154] ainsi que Zadeh et al. [155] ont utilisé des méthodes de  $K$  plus proches voisins pour la classification de microcalcifications. Une autre méthode de classification regroupe les réseaux de neurones [156, 157]. Le plus communément utilisé est le perceptron où les neurones. Dans le cas où le problème n'est pas linéairement séparable, plusieurs couches de neurones sont nécessaires. Ce type de réseaux a été utilisé pour la classification des microcalcifications [158, 159, 160, 161, 162, 163]. Une autre approche de classification est connue sous le nom de machines à vecteurs de support. Dans certains cas on emploie le terme de séparateurs à vaste marge Vapnik et al. [164]. L'idée est de trouver un hyperplan qui sépare deux classes en maximisant la marge de séparation des éléments les plus proches des deux classes. Bien entendu, souvent les classes ne sont pas linéairement séparables. Dans de tels cas, les éléments à classer sont transférés dans un espace de dimension supérieure où une séparation par un hyperplan aura plus de sens. Plusieurs travaux utilisent les SVM pour la classification Cao et al., Campanini et al. [165, 166].

### 4.1. Les types de la classification

La classification d'images est une application de la reconnaissance a pour but d'identifier ou attribuer automatiquement les classes auxquelles appartiennent des objets à partir de certains paramètres descriptifs et à l'aide d'un système de classification. Elle s'applique à un grand nombre d'activités comme la classification d'objets, de voix, la reconnaissance de visages, de textures, d'empreintes digitales, le diagnostic médicale [167]. Il existe deux types d'apprentissage pour les données à classer : l'apprentissage supervisé et l'apprentissage non-supervisé.

— La classification non supervisée : est une classification sans apprentissage qui consiste

à découper l'espace de représentation en zones homogènes selon un critère de vraisemblance entre les pixels. Cette approche est utilisée pour effectuer une classification de pixels, sans connaissance a priori sur l'image. Cette forme de classification existe depuis des temps immémoriaux. Elle concerne notamment les sciences de la nature (voir figure 3.2), les classifications des documents et des livres mais également la classification des sciences élaborées au cours des siècles par les philosophes [168].

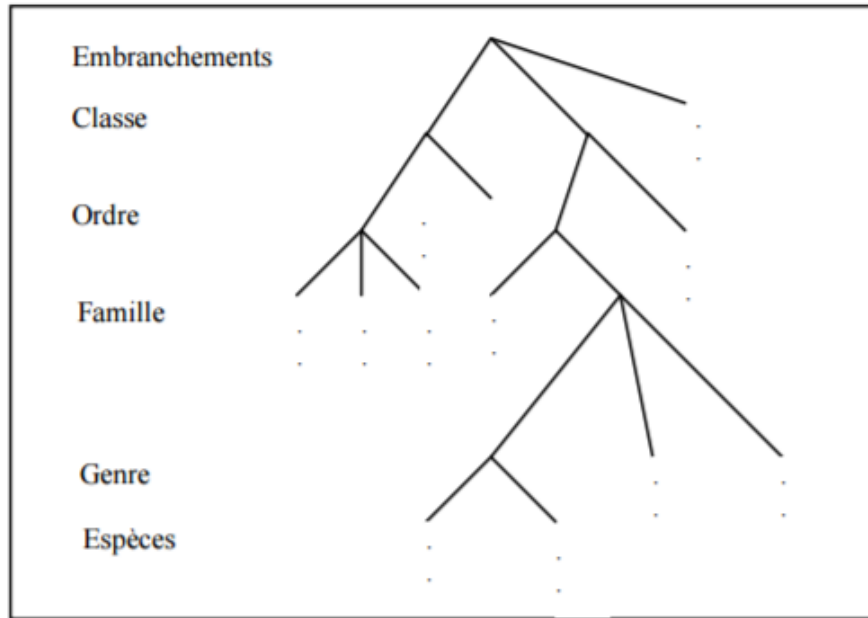


FIGURE 3.2 – Extrait de la classification taxinomique de Linné.

- La classification supervisée : est bien évidemment avec apprentissage, elle consiste à définir une fonction de discrimination effectuant une séparation dans l'espace de représentation à partir d'une connaissance des classes de l'image. Ce type de classification nécessite la création d'une base d'apprentissage faisant intervenir ce qu'on appelle une segmentation vérité terrain (voir figure 3.3).

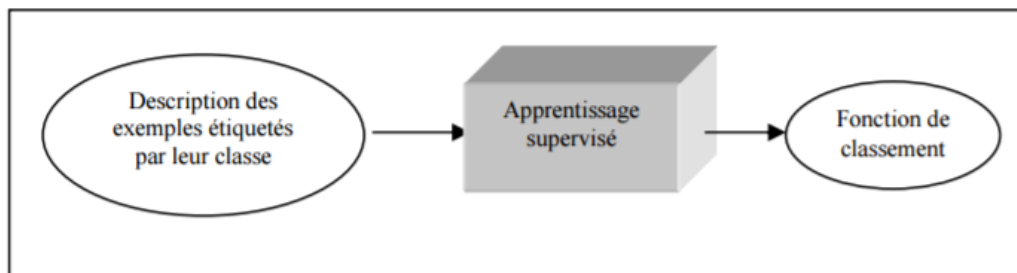


FIGURE 3.3 – L'apprentissage supervisé.

Les méthodes les utilisées pour la classification sont : la méthode de la Machine à Vecteurs

de Support (SVM), les Réseaux de Neurons (RN), K-Plus Proche voisin KPPV (en anglais k Nearest Neighbor KNN), etc. Nous présentons dans la suite de ce chapitre une étude détaillée des trois techniques SVM, KPPV et RN.

#### 4.1.1. Machine à Vecteurs de Support (SVM)

Les machines à vecteurs de support (en anglais Support Vector Machine, SVM) sont un ensemble de techniques de classification binaire par apprentissage supervisé. Les SVM sont une généralisation des classifieurs linéaires.

Les SVM ont été proposées par V. Vapnik en 1995 [169], ont rapidement été adoptés pour leur capacité à travailler avec des données de grandes dimensions et ont pu montrer l'efficacité de ces techniques dans nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finance...). Le principe de méthode SVM est de classer les données en deux classes, afin que les données non linéairement séparables dans l'espace d'entrée deviennent linéairement séparables dans l'espace de caractéristiques. Elle est basée sur l'existence d'un classificateur linéaire appelé hyperplan (voir figure 3.5) qui sépare les deux ensembles de points dans un espace approprié et sur l'utilisation de fonction dites noyau qui permet une séparation optimale des données de ces deux classes.

Parmi les modèles des SVM, on trouve les cas linéairement séparable( voir figure 3.4.a) et les cas non linéairement séparable( voir figure 3.4.b).

a) Cas linéairement séparable : plus simple pour trouver facilement le classificateur linéaire

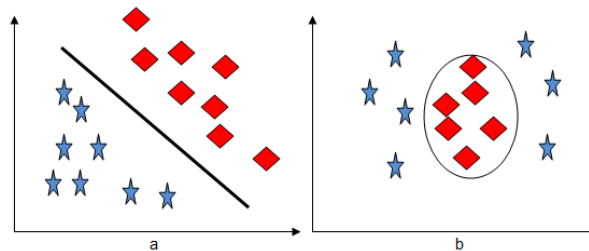


FIGURE 3.4 – Les modèles des SVM : a) cas linéairement séparable, b) non linéairement séparable.

Pour déterminer l'hyperplan séparable idéal on utilise des points qui sont appelés vecteurs supports, se sont les points les plus proches ou les points de la frontière entre les deux classes des données [170, 171] et une distance maximal qui s'appel la marge entre l'hyperplan et les vecteurs supports ( voir figure 3.5).

L'hyperplans à marge maximale a été explorée par Vladimir Vapnik qui est un séparateur qui passe au milieu des points des deux classes d'exemples. L'équation de cet hyperplans est donnée par :

$$W.x + b = 0 \tag{3.28}$$

A la présence d'un ensemble d'exemples  $(x_i, y_i)$  Le but est donc de :

- Maximiser la marge :  $2/\|W\|$

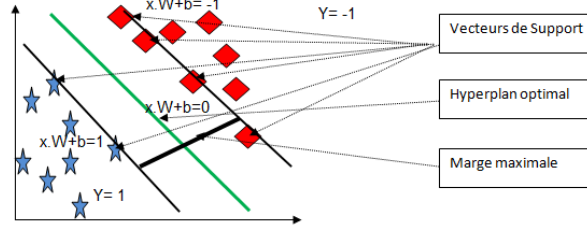


FIGURE 3.5 – L’hyperplan H optimal, vecteurs supports et marge maximale.

- Garder une séparation correcte, autrement dit :  
 Pour les  $x$  positifs ( $y=1$ ) :  $W.x + b \geq 1$   
 Pour les  $x$  négatifs ( $y=-1$ ) :  $W.x + b \leq -1$

Ces conditions peuvent être formulées sous forme du problème d’optimisation quadratique suivant :

$$\begin{cases} \min_{W,b} (\frac{\|W\|^2}{2}) \\ y_i(W.x_i + b) \geq 1 \end{cases}$$

Il s’agit d’un problème quadratique dont la fonction objective est à minimiser, pour cela il est nécessaire d’introduire une formulation dite duale, alors pour dualiser, nous devons former ce que l’on appelle le Lagrangien. Il s’agit de faire rentrer les contraintes dans la fonction objective et de pondérer chacune d’entre elles par une variable duale :

$$\partial_\omega(\omega, b, \alpha) = \omega - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (3.29)$$

$$\partial_b(\omega, b, \alpha) = \sum_{i=1}^n \alpha_i y_i = 0 \quad (3.30)$$

$$\partial_{\alpha_i}(\omega, b, \alpha) = y_i(\omega_i . x_i + b) + 1 \leq 0 \quad (3.31)$$

$$\alpha_i = y_i(\omega_i . x_i + b) - 1 = 0 \forall i = 1 \dots n \quad (3.32)$$

L’équation (3.29), permet de réexprimer  $w$  :

$$\omega = \sum_{i=1}^n \alpha_i y_i x_i \quad (3.33)$$

nous substituons l’équation (3.30) et l’équation (3.33) dans le Lagrangien :

$$L(\omega, b, \alpha) = \frac{1}{2} \langle \omega, \omega \rangle - \sum_{i=1}^n \alpha_i [y_i (\langle \omega, x_i \rangle + b) - 1] \quad (3.34)$$

$$= \frac{1}{2} \sum_{ij=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle - \sum_{ij=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle + \sum_{i=0}^n \alpha_i \quad (3.35)$$

$$= -\frac{1}{2} \sum_{ij=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle + \sum_{i=0}^n \alpha_i \quad (3.36)$$

maximiser :

$$W(\alpha) = -\frac{1}{2} \sum_{ij=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle + \sum_{i=0}^n \alpha_i \quad (3.37)$$

tel que

$$\begin{cases} \sum_{i=0}^n \alpha_i y_i = 0 \\ \alpha_i \geq 0 \forall i = 1 \dots n \end{cases}$$

la fonction de décision :

$$f(x) = \sum_{i=1}^n \alpha_i y_i < x, x_i > + b \quad (3.38)$$

b) Cas linéairement non séparable :

si il n'y a pas de séparation lineaire entre les données dans cas le classifieur de marge maximale ne peut pas être utilisé. Pour resoudre cette nouvelle contrainte, nous allons utilisé une technique dite des variables ressort (slack variables).

Le probleme d'optimisation initial était à pour but de minimiser  $\frac{1}{2}\|W\|^2$  tel que  $y_i(\omega x_i + b) \geq 1$ . les contraintes sont donc relâchées et qui deviennent :

si  $y_i = +1$

$$\omega x_i + b \geq 1 - \xi_i \quad (3.39)$$

si  $y_i = -1$

$$\omega x_i + b \leq -1 + \xi_i \quad (3.40)$$

ce qui revient à :

$$y_i(\omega x_i + b) \geq 1 - \xi_i \forall i = 1 \dots n. \quad (3.41)$$

si  $\xi$  est supérieur à 1, le classifieur attribuera une classe fausse à un element  $x_i$ , donc  $\sum_i \xi_i$  est une borne supérieure du nombre d'erreurs à l'apprentissage. Dans ce cas non-linéaire, pour résoudre le problème est de remplacer la fonction précédente à minimiser par :

$$J = \frac{1}{2}\|W\|^2 + C \sum_{i=1}^n \xi_i \quad (3.42)$$

Le coefficient C est défini comme un paramètre de régularisation, il donne un compromis entre la marge et le nombre d'erreurs. Une faible valeur de C entraîne une faible tolérance. Le Lagragien dans ce cas est donné par :

$$L_p(\omega, b, \xi, \alpha, \mu) = J(\omega, \xi) + \sum_{i=1}^n [1 - \xi_i - y_i(\omega x_i + b)] \mu_i - \sum_{i=1}^n \xi_i \mu_i \quad (3.43)$$

La resolution du probleme d'optimisation est similaire que le premier cas, il faut trouver de *même* une solution  $\alpha$ . aux conditions de Karush-Kuhn-Tucker (KKT) [172].

c) Cas non linéaire :

Le classificateur de marge maximale ne peut pas être utilisé dans ce cas, car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables, alors l'SVM permet de résoudre ce problème et retrouver le cas linéairement séparable par le changement de l'espace des données qui s'appelle espace de redescription ( voir figure 3.6), cette transformation non linéaire est réalisée via une fonction noyau. Dans ce nouvel espace, l'équation de l'hyperplan est donnée par :

$$\sum_i \alpha_i y_i \varphi(x_i) \cdot \varphi(x) + b = 0 \quad (3.44)$$

Au lieu de définir la transformation, il suffit de choisir une fonction  $K(x, y)$  et supposer qu'il existe une transformation  $\varphi$  telle que :

$$k(x, y) = \varphi(x) \cdot \varphi(y) \quad (3.45)$$

Cette fonction s'appelle un noyau (Kernel). On n'est pas alors supposé connaître le nouvel espace, on s'intéresse uniquement au produit scalaire dans cet espace représenté par le Noyau K. Pour qu'une fonction  $K(x, y)$  soit un noyau valide, elle doit vérifier certaines conditions, nommées les conditions de Mercer.

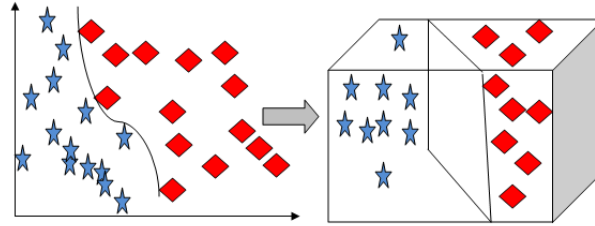


FIGURE 3.6 – Transformation des données dans un espace de grande dimension .

#### 4.1.2. K Plus Proches Voisins (KPPV)

La méthode de K Plus Proches Voisins KPPV (en anglais k Nearest Neighbor) parmi les plus simples algorithmes d'apprentissage artificiel supervisé, introduite par Cover et Hart [173].

l'idée de k plus proches voisins est simple : dans la partie de l'apprentissage on a une base d'apprentissage qui contient des pixels classés soit à disposition. Une donnée de classe inconnue à classer est comparé aux autres données stockées de la base d'apprentissage ayant déjà été classés, et on lui affecte la classe la plus représentée parmi les k-plus proche à l'aide d'une mesure de similarité. Pour une donnée ou (le candidat A ou B) qui doit être assigné à l'un des deux groupes, les k voisins les plus proches parmi les candidats de l'ensemble d'apprentissage sont déterminés, et les données ou le candidat est assigné au groupe auquel la majorité parmi les k voisins appartient ( voir figure 3.7).

le choix de la valeur de k et la fonction de distance pertinente posent des problèmes malgré la facilité de cette technique et les bons résultats à donner. Le choix de la valeur de k est évalué expérimentalement en fonction du taux de bonne classification avec la distance euclidienne.

— La distance euclidienne [174] :

La distance euclidienne entre deux points, x et y, avec n dimensions est calculé comme suit :

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.46)$$

La distance euclidienne est toujours supérieure ou égale à zéro. La mesure serait de zéro pour les points identiques et haute pour les points qui montrent peu de similitudes.

— La distance City Block [174] :

$$d = \sum_{i=1}^n |x_i - y_i| \quad (3.47)$$

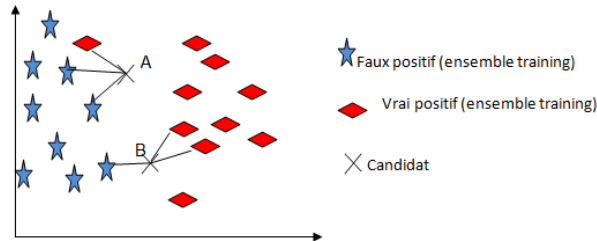
Dans la plupart des cas, cette mesure de distance donne des résultats similaires à la distance euclidienne. Le nom de la distance City block (également dénommé distance de Manhattan) est expliqué si l'on considère deux points dans le plan xy. La plus courte distance entre les deux points est le long de l'hypoténuse, qui est la distance euclidienne. La distance de bloc city est la place calculé comme la distance en x plus la distance en y, qui est similaire à la façon dont vous déplacez dans une city (comme Manhattan).



— La distance Minkowski [174] :

$$d = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (3.48)$$

Pour  $p=1$ , on va avoir la distance de City Block, et pour  $p=2$  on va avoir la distance euclidienne




---

FIGURE 3.7 – Principe de classification par KPPV

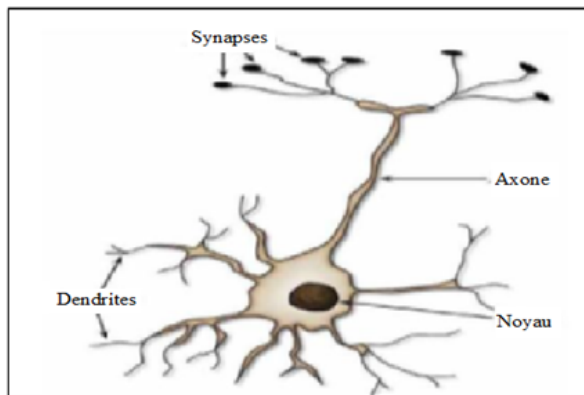
#### 4.1.3. Réseaux de Neurones (RN)

Un réseau de neurones artificiels est une conception de calcul inspirée du fonctionnement des neurones biologiques (voir figure 3.8). Haykin et al. ont proposé la définition suivante [175] :

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 :

1. La connaissance est acquise à travers d'un processus d'apprentissage.
2. Les poids des connections entre les neurones sont utilisés pour mémoriser la connaissance.

Sur base de cette définition, repose l'élaboration des réseaux de neurones artificiels.




---

FIGURE 3.8 – Représentation de la forme biologique d'un neurone .

Le neurone est constitué par quatre unités : le soma, les axones, les dendrites et les synapses. Chacune de ces différentes parties joue un rôle bien défini dans le transport de l'information.

— Les dendrites :

Les dendrites sont des extensions, des ramifications tubulaires, ce sont les entrées principales du neurone qui permettent d'acheminer des informations en provenance d'autres neurones et transmettent au corps du neurone.

— Le soma :

Le soma, qui contient la partie centrale d'un neurone, car il traite les signaux qu'ils reçoivent des dendrites pour transmettre les informations utiles aux axones.

— L'axone :

L'axone est un prolongement non ramifié du soma. Sert à transmettre les informations électrique et chimique par cellules du cerveau vers d'autres neurones qui l'entourent.

— Les synapses :

Une interface entre l'axone du neurone et les dendrites des autres neurones.

Pour étudier le neurone artificiel, le neuropsychiatre McCulloch et le logicien Pitts [176] s'inspirant de leurs travaux sur les neurones biologiques, proposèrent en 1943 le modèle suivant ( voir figure 3.9) :

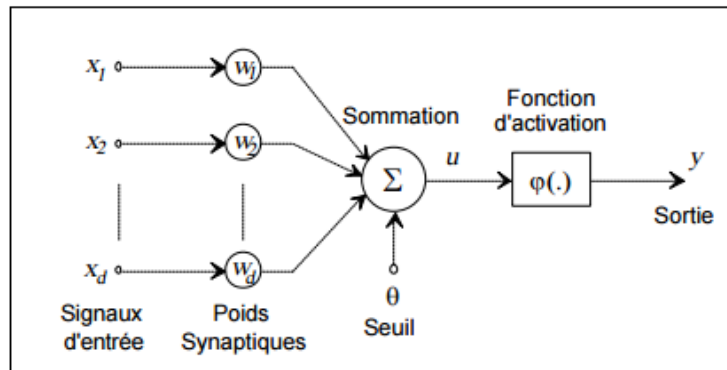


FIGURE 3.9 – Le modèle de neurone formel

Le neurone formel est un modèle qui se caractérise par les entrées qui sont désignées par  $x_1, x_2, x_3, \dots, x_d$  et une fonction de transition d'état .

$$y = f\left(\sum_{i=1}^d W_i x_i\right) \quad (3.49)$$

Le neurone réalise alors trois opérations sur ses entrées :

— Pondération :

multiplication de chaque entrée par un paramètre appelé poids de connexion.

— Somme :

une somme des entrées pondérées est effectuée

— Activation :

passage de cette somme dans une fonction, appelée fonction d'activation.

La valeur calculée a la sortie du neurone qui est transmise aux neurones suivants.

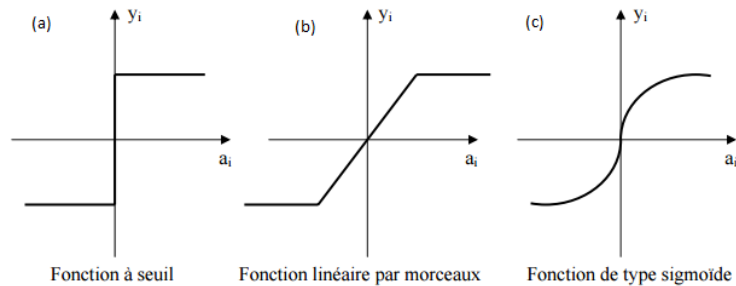


FIGURE 3.10 – Les fonction de transfert pour le neurone artificiel.

— La fonction seuil :

On applique un seuil sur son entrée. Cette fonction permet de prendre des décisions binaires (vraie ou faux) ; la fonction retourne la valeur 0 si l'entrée est négative sinon ils dépassant le seuil pour une entrée positive ou nulle, la fonction retourne 1. Cette fonction est illustrée dans la figure 3.10.(a).

— La fonction lineaire :

elle affecte directement son entrée à sa sortie. Cette fonction est illustrée dans la figure 3.10.(b).

— La fonction sigmoïde :

elle prend ses valeurs entre 0 et 1, cette fonction est illustrée à la figure 3.10.(c).

### 1. Perceptron monocouche

Le perceptron est considéré comme le premier modèle des réseaux de neurones, il fut mis au point dans les années cinquante par Rosenblatt (1957-1961) [177]. Le perceptron se compose de deux couches de neurones la rétine (n'est pas compté d'où le nom de perception monocouche) et la couche de sortie. La fonction seuil de Heaviside est utilisée comme fonction d'activation des neurones de la couche de sortie. La figure (3.11) montre un exemple de perceptron.

Les cellules de la première couche sont binaires, répondent en oui / non (0/1). Les cellules d'entrée sont reliées aux neurones de sortie grâce à des liens synaptiques  $W_{ij}$  d'intensité variable. La règle d'apprentissage du perceptron est la règle de Widrow-Hoff [178] :

$$W_{ij}^{t+1} = W_{ij}^t + \eta(t_j - o_j)x_i = W_{ij}^t + \Delta W_{ij} \quad (3.50)$$

Avec :

- $\Delta W_{ij}$  : Changement à effectuer pour la valeur  $W_{ij}$ .
- $x_i$  : Valeur de sortie ( 0 ou 1 ) de la  $i^{\text{ème}}$  cellule de la rétine.
- $o_j$  : Réponse de la  $j^{\text{ème}}$  cellule de sortie ( 0 ou 1 ).
- $t_j$  : Réponse théorique ou (désirée) de la  $j^{\text{ème}}$  cellule de sortie (0 ou 1).
- $W_{ij}^t$  : Intensité de la connexion entre la  $i^{\text{ème}}$  cellule d'entrée et la  $j^{\text{ème}}$  cellule de sortie, au temps t (les valeurs  $w_{ij}^0$ ) et  $w_0$  sont généralement choisies d'une manière aléatoire.

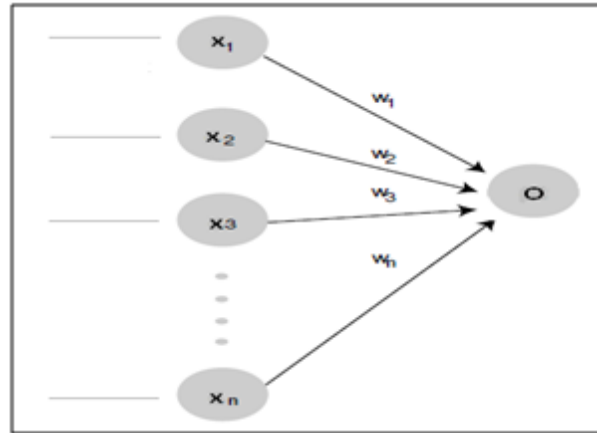


FIGURE 3.11 – Exemple d'un perceptron monocouche

- $\eta$  : une constante positive généralement comprise entre 0 et 1, sa valeur influe, en effet, sur la vitesse d'apprentissage.

Ce type de perceptron est limité à effectuer la classification dans un problème à deux classes seulement. La reconnaissance de plusieurs classes est cependant rendue possible par la mise en parallèle de plusieurs perceptrons. Le perceptron ainsi obtenu comporte un neurone par classe, chacun de ceux-ci réalisant une fonction discriminante linéaire de la classe à laquelle il est associé. Ainsi, les frontières de décision pour la classification offertes par le perceptron sont linéaires. Néanmoins, de nombreux problèmes de classification nécessitent une partition non linéaire de l'espace d'entrée. Ces problèmes peuvent être résolus à l'aide des perceptrons multicouches, qui consistent à mettre en cascade deux couches ou plus de perceptrons.

- Perceptron multicouche

Les réseaux à deux couches peuvent être entraînés avec des règles d'apprentissage relativement simples. En contrepartie les réseaux à deux couches sont limités au calcul de fonctions très simples. De là, l'intérêt d'utiliser des réseaux plus évolués, contenant en outre des neurones cachés. Le perceptron multicouches peut se voir comme une généralisation du perceptron, il se compose d'une première couche de cellules d'entrée (équivalente à la rétine du perceptron), d'une ou plusieurs couches intermédiaires (dites couches cachées), et d'une couche de sortie qui est la seule en contact avec le monde extérieur (ou le superviseur).

La connectivité de ce réseau est restreinte : un neurone d'une couche inférieure ne peut être relié qu'à des neurones de couche suivante. Le calcul de l'activation dans le réseau s'effectue en propageant l'activation initiale de la couche d'entrée jusqu'à la couche de sortie. L'erreur est calculée dans le sens inverse. En général, les neurones du PMC sont animés par une fonction d'activation sigmoïde, soit exponentielle, soit tangentielle. La non linéarité de ces fonctions est nécessaire pour que les couches cachées du réseau soient utiles. De plus, leur dérivabilité est nécessaire à l'apprentissage du réseau. La figure( 3.12) montre un perceptron multicouche composé d'une couche d'entrée, une seule couche caché et une couche de sortie.

Pour répondre à un stimulus, le signal est propagé de la couche d'entrée à la couche de sortie en passant par la couche cachée, ainsi lorsque un stimulus est présenté en entrée, la réponse du  $m^{\text{ième}}$  neurone de la couche cachée  $h_m$  est donné par :

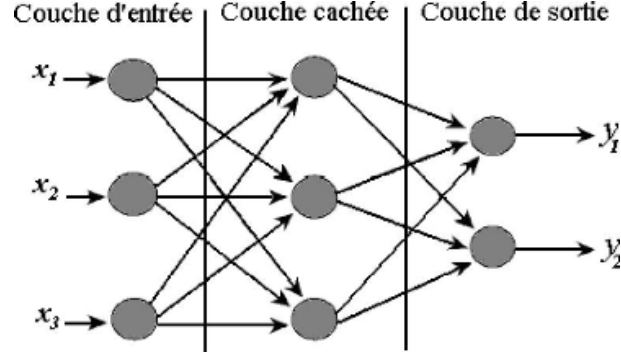


FIGURE 3.12 – Perceptron multicouches avec une seule couche cachée

$$h_m = f(\sum_{i=1}^I w_{0m} + x_i w_{im}) = f(\sum_{i=0}^I x_i w_{im}) \quad (3.51)$$

ou  $x_0 = 1$

$w_{0m}$  est le neurone d'entrée correspondant au biais. Puis la réponse du  $j^{\text{ième}}$  neurone de la couche de sortie  $o_j$  est donnée par :

$$o_j = f(z_{0j} + \sum_{m=1}^M h_m + z_{mj} w_{im}) = f(\sum_{m=0}^M h_m + z_{mj} w_{im}) \quad (3.52)$$

ou  $h_0 = 1$

$z_{0j}$  est le neurone de la couche cachée correspondant au biais.  $f$  est la fonction d'activation sigmoïde :

$$f(a) = \frac{1}{1 + e^{-a}} \quad (3.53)$$

La règle d'apprentissage utilisée est la rétro-propagation de l'erreur, qui est une technique de descente du gradient minimisant l'erreur instantanée entre la sortie du réseau  $o^k$  et celle désirée  $t^k$  lorsque le  $k^{\text{ième}}$  stimulus est présenté en entrée :

$$E_k = \frac{1}{2} \sum_{j=1}^J (t_j^k - o_j^k)^2 \quad (3.54)$$

L'adaptation des poids synaptiques se calcule de manière itérative pour tous les  $K$  stimulus jusqu'au moment où un minimum de l'erreur est trouvé. La fin de l'apprentissage peut être décidée lorsque l'erreur totale est

$$E_T = \sum_{k=1}^K E_k \quad (3.55)$$

Observée à la sortie du réseau devient inférieure à un seuil prédéfini. La règle d'adaptation avec l'ajout du moment de Rumelhart est définie comme suit. Calculer le signal d'erreur pour les neurones de la couche de sortie et la couche cachée respectivement :

$$\delta_j^k = (t_j^k - o_j^k) o_j^k (1 - o_j^k) \quad (3.56)$$

$$\delta_m^k = h_m^k (1 - h_m^k) \sum_{j=0}^J \delta_j^k z_{mj} \quad (3.57)$$

Corriger l'ensemble des poids synaptique  $z_{mj}$  et  $w_{im}$  suivant les formules suivantes :

$$z_{mj}(t + 1) = z_{mj}(t) + \eta \delta_j^k h_m^k + \mu(z_{mj}(t) - z_{mj}(t - 1)) \quad (3.58)$$

$$w_{im}(t + 1) = w_{im}(t) + \eta \delta_m^k x_i^k + \mu(w_{im}(t) - w_{im}(t - 1)) \quad (3.59)$$

Avec  $\eta$  et  $\mu$  sont respectivement le pas d'apprentissage et le moment. Pour une tâche de classification, la couche de sortie comporte un neurone par classe et les valeurs désirées sont : 1 pour le neurone qui correspond à la classe de l'objet présenté à l'entrée et 0 pour tous les autres neurones. De la définition de la fonction d'erreur, l'apprentissage tend à augmenter la valeur de sortie du neurone qui correspond à la classe correcte, tout en cherchant à diminuer les valeurs des sorties des autres neurones. Il en résulte donc un apprentissage discriminant. En pratique cependant, même lorsque seule la fonction sigmoïde est utilisée, la somme des sorties des neurones possède souvent une tendance naturelle à s'approcher aussi nettement de l'unité.

## 5. Evaluation de la performance de classification

Considérons un échantillon de sujets extrait au hasard d'une population. Les sujets sont classés au malade (M+) ou non malade (M-) à l'aide d'une méthode dite de référence ayant fait la preuve de sa valeur diagnostique. Les vrais positifs (VP) sont les résultats positifs chez les sujets porteurs de la maladie, les faux positifs (FP) sont les résultats positifs chez les sujets indemnes de la maladie. De même, les vrais négatifs (VN) sont les résultats négatifs chez les sujets non malades et les faux négatifs (FN) sont les résultats négatifs chez les sujets malades tableau (3.1).

Tableau 3.1 – Signification de VP, VN, FP et FN de la matrice de confusion

Mesure du rendement	Définition
VP(Vrais Positifs)	classement correct des anormales
FP(Faux Positifs)	classement incorrect des anormales
FN(Faux Négatifs)	classement incorrect des normales
VN(Vrais Négatifs)	classement correct des normales

On résume les résultats d'un classifieur sur des données de test pour chaque classe dans une matrice dite : matrice de confusion comme le montre le tableau (3.2).

Tableau 3.2 – Matrice de confusion

Réponse du classifieur	Réponse de l'expert	
	Positifs	Négatifs
Positifs	VP(Vrais Positifs)	FP(Faux Positifs)
Négatifs	FN(Faux Négatifs)	VN(Vrais Négatifs)

Afin d'évaluer les performances des systèmes de classification nous avons calculer cinq mesures de la performance qui sont [179, 180] :

- La sensibilité est définie comme étant le rapport des tumeurs qui sont marqués et classé comme une tumeur, à toutes les tumeurs marquées, donnée par :

$$Se = \frac{VP}{VP + FN} \quad (3.60)$$

- La spécificité est définie comme le rapport des tumeurs qui ne sont pas marqués et également pas classés comme tumeur, de toutes les tumeurs non marquées, donnée par :

$$Sp = \frac{VN}{VN + FP} \quad (3.61)$$

On résume La sensibilité et la spécificité du test sont estimées respectivement par la proportion de vrais positifs chez les malades et la proportion de vrais négatifs chez les non malades [179].

Outre les paramètres évoqués précédemment (Se, Sp), il en existe d'autres bien connus et largement utilisés pour évaluer les performances des classifieurs, et qui peuvent être calculés à partir de la matrice de confusion comme suit :

- La fraction positive des indicateurs de performance vrai (VPF) et la fraction de faux positifs (FPF) peut être calculé en utilisant :

$$FPR = \frac{FP}{VN + FP} = 1 - Sp \quad (3.62)$$

$$VPR = \frac{VP}{VP + FN} = Se \quad (3.63)$$

- L'exactitude globale est le rapport entre le nombre total de cas correctement classés et la taille de l'ensemble de test, donnée par :

$$Exactitude = \frac{VP + VN}{VP + FN + VN + FP} \quad (3.64)$$

$$Précision = \frac{VP}{VP + FP} \quad (3.65)$$

Le taux de équilibre d'un classification(BCR :Balanced Classification Rate) est une métrique d'évaluation couramment utilisés pour les ensembles de données déséquilibrées. Depuis BCR fournit une moyenne de sensibilité et de spécificité, il donne une image plus précise de l'efficacité du classificateur.

$$BCR = \frac{Se + Sp}{2} \quad (3.66)$$

Le taux d'erreur équilibrée (BER :Balanced Error Rate) .

$$BER = 1 - BCR \quad (3.67)$$

## Conclusion

Dans ce chapitre, nous avons établi un état de l'art des différentes méthodes qui abordent le problème de l'apprentissage dans notre domaine d'intérêt : la reconnaissance des images mammographique, par la présentation des différents descripteurs et classificateurs pour l'extraction des caractéristiques de texture à savoir les matrices de co-occurrence, les attributs de Tamura, Gabor, Gist, et les ondelettes. Aussi nous avons présenté dans ce chapitre les différentes méthodes de classification que nous avons utilisé pour classer les images tel que les réseaux de neurones RN, les machines à vecteurs de support SVM, et les k-plus proche voisin KPPV.



# Chapitre 4

## Contributions : système de segmentation et classification des images mammographiques

### Introduction

L'objectif de cette thèse est de fournir un système d'aide au diagnostic précoce de la présence du cancer du sein ou des microcalcifications dans les images mammographiques. La figure (4.1) montrent les différentes étapes pour l'élaboration de ce système. Ainsi, dans ce chapitre, nous présentons nos différentes contributions correspondantes aux trois phases du système proposé. Dans la première section, nous présentons la technique de segmentation du muscle pectoral afin d'avoir une image prête à être traitée. Cette image résultat subira par la suite un ensemble de techniques de rehaussement du contraste afin d'améliorer la qualité des images mammographiques. La dernière section représente notre dernière contribution par rapport à l'extraction des caractéristiques et la classification des images mammographiques. Les résultats obtenus et l'évaluation des performances des algorithmes sont également présentés.

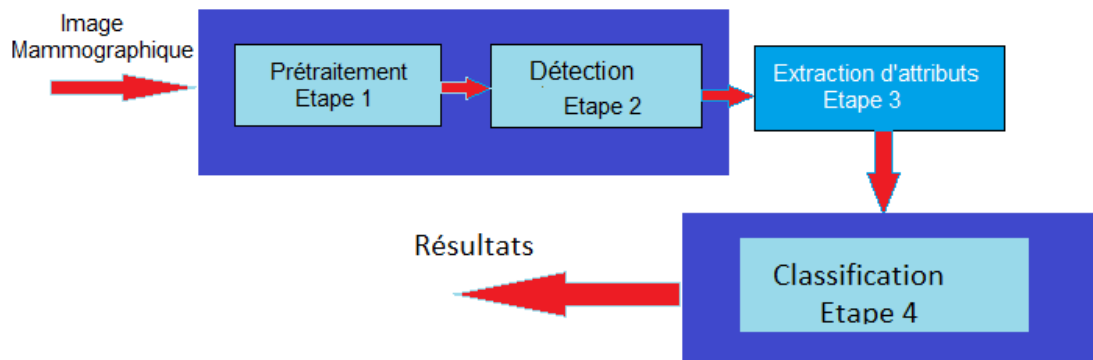


FIGURE 4.1 – Système de travail

## 1. Segmentation du muscle pectoral

Dans les images mammographiques, le muscle pectoral est un tissu plus dense que ceux formant le sein et apparaît comme une région triangulaire de haute intensité dans la marge supérieure. Selon l'anatomie de la patiente et son positionnement durant l'acquisition, le muscle pectoral peut occuper plus que la moitié de la région du sein ou juste une petite portion de l'image. La frontière entre le muscle pectoral et les tissus fibreux étant floue et irrégulière à cause du bruit dans le cliché, il existe alors un gradient en niveau de gris difficile à analyser [112]. En général, la courbure du muscle pectoral est convexe et peut dans certains cas être concave ou les deux à la fois. Dans la plupart des cas, la partie supérieure du muscle a un contour très clair avec une intensité tranchante ; par contre la partie inférieure présente un contour plus texturé du fait que cette partie se chevauche avec la matière fibreuse [57].

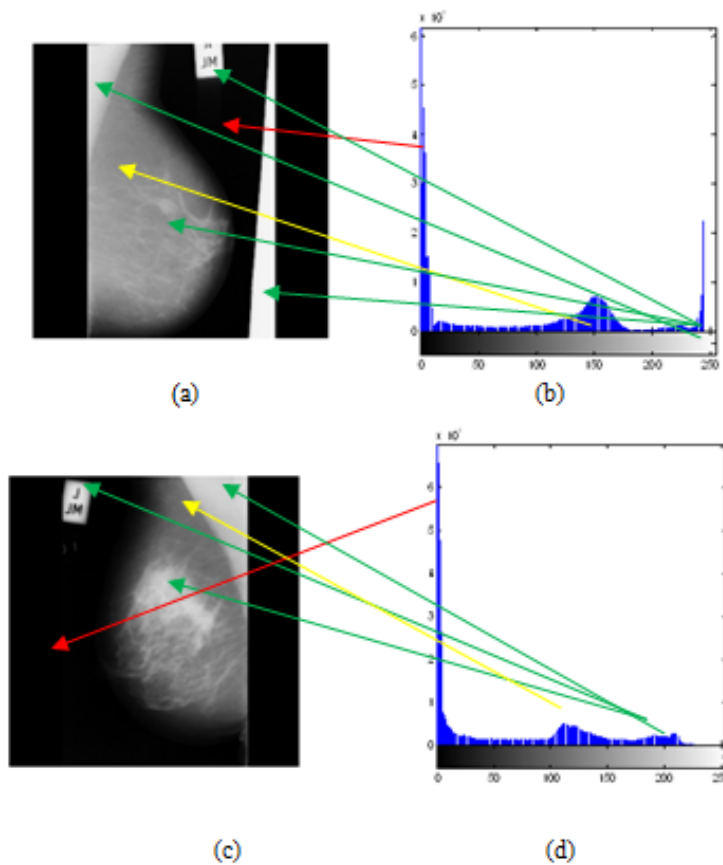


FIGURE 4.2 – Images mammographiques : a) image (mdb012), b) son histogramme, c) image (mdb209), d) son histogramme

À partir de ces histogrammes, nous constatons la présence de trois classes différentes :

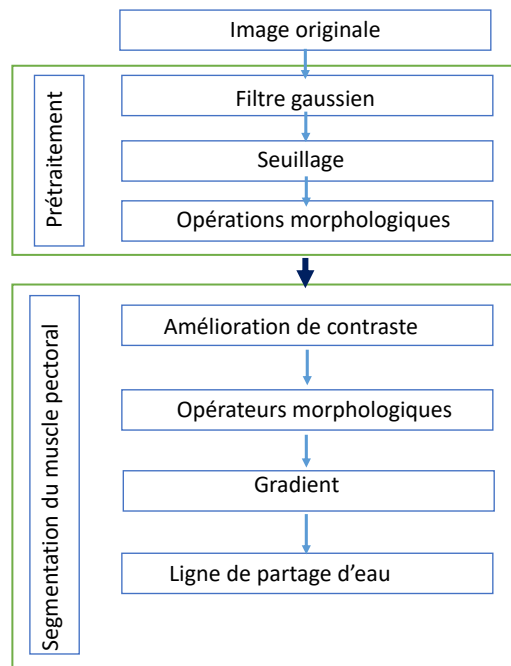
- Classe 1 : elle correspond aux pixels de faible intensité (la gauche des histogrammes) représentant le fond des images.
- Classe 2 : elle correspond aux pixels d'intensité moyen de gris représentant le sein (

ou poitrine).

- Classe 3 : elle correspond à tous les pixels de forte intensité et qui représentent le muscle pectoral, des artefacts de radio-opaque, et peut être les lésions mammaires.

À partir de ces constatations, il se montre nécessaire et utile de supprimer la classe 3 qui correspond notamment au muscle pectoral ainsi que les artefacts, etc, afin d’avoir une image claire du sein prête à être traitée. Pour cela, nous proposons dans la figure (4.3) notre démarche pour nettoyer les images mammographiques brutes. Notre approche se déroule en deux étapes :

1. Phase de prétraitement qui consiste à supprimer tout bruit altérant la qualité des images.
2. Phase de segmentation qui consiste à détecter le muscle pectoral et de supprimer des images mammographiques.



---

FIGURE 4.3 – Schéma proposé de segmentation du muscle pectoral

### 1.1. Prétraitement de l'image

La phase de prétraitement est une phase primordiale dans tous processus de traitement d'images.

Dans notre cas, les images mammographiques obtenues sont altérées par le bruit qui affecte l'analyse et une prise de décision correcte ; ce qui nous oblige dans un premier temps à supprimer le bruit. Le bruit apparaissant dans les images mammographiques est le bruit de numérisation, les artefacts et les muscles pectoraux.

### 1.1.1. Suppression du bruit

Les bruits de numérisation tels que les lignes verticales et horizontales qui apparaissent sur la plupart des images de mammographie, tel qu'il figuré par les marqueurs d'indication sur la figure (4.4). Ces bruits sont supprimés par le filtre Gaussien présenté suivant l'équation (4.2) avec l'écart type  $\sigma = 6$  [181].

$$h_g(n_1, n_2) = \exp \frac{-(n_1^2+n_2^2)}{2\sigma^2} \quad (4.1)$$

$$h(n_1, n_2) = \frac{h_g(n_1, n_2)}{\sum_{n_1} \sum_{n_2} h_g} \quad (4.2)$$

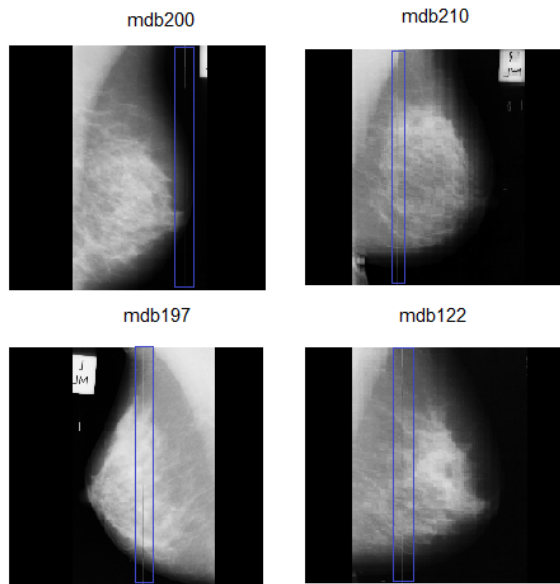


FIGURE 4.4 – Bruits de numérisation en images mammographiques

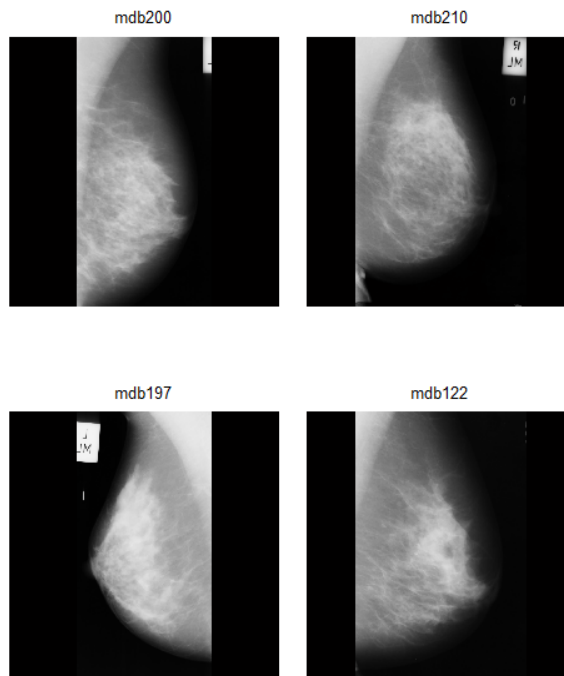
La figure (4.5) montre les images de mammographie résultantes après l'application du filtre Gaussien. La suppression du bruit de numérisation dans les images de mammographie n'affecte pas le profil du sein.

### 1.1.2. Suppression des artefacts Radiopaques

Les Radio-opaques, tels que les coins et les étiquettes dans les images de mammographie (voir figure 4.6) peuvent contribuer à incrémenter les difficultés pour les algorithmes de segmentation et d'amélioration [182].

La démarche pour supprimer les artefacts est comme suit :

1. Segmenter les objets et l'arrière-plan, en binarisant l'image mammographique par la technique du seuillage avec un seuil ( $T=0.13$ ). L'image résultat après le seuillage est présentée sur la figure (4.7.a). À l'issue du seuillage, nous récupérons plusieurs composantes de tailles différentes qui correspondent au sein et aux artefacts qu'ils faut éliminer.



---

FIGURE 4.5 – Images de mammographie après suppression du bruit par le filtre Gaussien

2. Ensuite, on procède par des opérations de fermeture morphologiques pour délimiter les artefacts à supprimer. L'image résultat à cette étape est illustrée sur la figure (4.7.b).
3. L'image binaire obtenue à l'étape (2) ( voir figure 4.7.b) est multipliée par l'image mammographique initiale pour obtenir une image claire sans bruit ( voir figure 4.8).

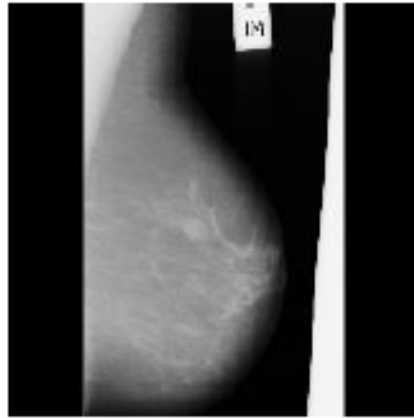
## 1.2. Suppression des muscles pectoraux

Après la phase de prétraitement, où on a supprimé le bruit ainsi que les artefacts, l'image obtenue doit subir un autre traitement qui concerne la suppression du muscle pectoral.

La suppression des muscles pectoraux est très complexe dû, d'une part, de la diversité de leurs formes et, d'autre part, de la frontière mal définie entre le tissu sein et la zone du muscle pectoral. Tenant compte de ces difficultés, nous avons proposé une approche de segmentation du muscle pectoral basée sur la technique de segmentation : la Ligne de Partages des Eaux (LPE) contrôlée par marqueurs [183]. Le principe de cet algorithme est de considérer l'image comme un plan géographique où chaque niveau de gris est interprété comme une altitude en relief topographique. Les pixels sombres correspondent aux vallées tandis que les pixels lumineux correspondent à des collines et des lignes de crête. La démarche proposée pour la segmentation suit les étapes suivantes :

### 1. Filtrage par le filtre médian

Ce filtre permet de réduire la présence du bruit réparti de façon aléatoire sur l'image. Ce type de filtrage préserve les contours des objets, ce qui est souhaitable. L'image résultante est présentée sur la figure (4.9.a)



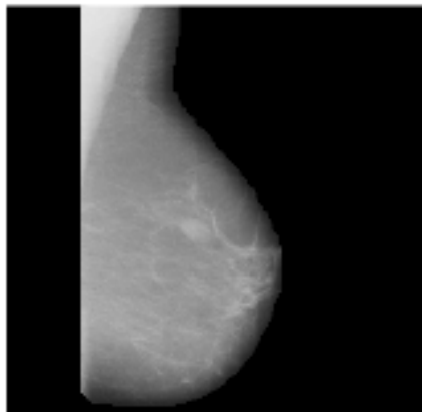
---

FIGURE 4.6 – Image (mdb012) mammographique initiale



---

FIGURE 4.7 – Images résultantes avec : a) seuillage  $T = 0.13$ , b) puis opérations morphologiques  $R=55$



---

FIGURE 4.8 – Image mammographique après prétraitement

## 2. Amélioration du contraste

Afin de bien cerner le muscle pectoral, le contraste du profil du sein, en particulier la région du muscle, doit être bien renforcée. Cela se fait par une méthode standard de réhaussement du contraste. L'image résultante est présentée sur la figure (4.9.b).

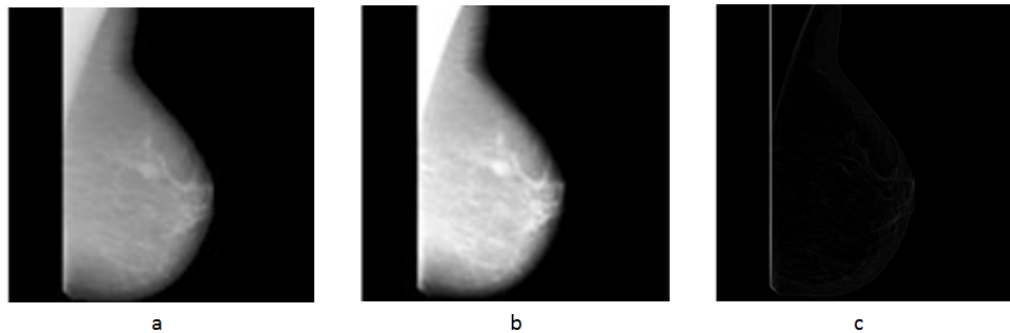


FIGURE 4.9 – Principales étapes de segmentation du muscle pectoral : a) Image originale, b) prétraitement, c) gradient

3. **L'application direct de la LPE sur l'image gradient** donne une sur-segmentation. Un filtrage morphologique est nécessaire pour réduire le nombre des minima élevés. L'image résultante est présentée sur la figure (4.10.a).
4. **Opérateurs morphologiques**

La morphologie mathématique est appliquée ici pour trouver les marqueurs du premier plan, qui doivent être connectés à l'intérieur de chacun des objets. Nous allons utiliser les techniques morphologiques d'ouverture par reconstruction et de fermeture par reconstruction pour nettoyer l'image et réduire le nombre des minima (voir figure 4.10.b). Ces opérations vont créer un maxima plat à l'intérieur de chaque objet pour trouver les marqueurs du premier plan comme montrée sur la figure (4.10.c).

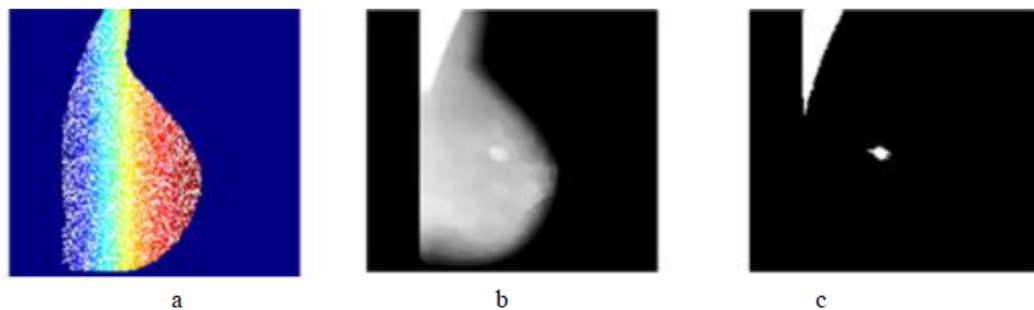


FIGURE 4.10 – Principales étapes de segmentation du muscle pectoral : a) LPE, b) morphologie mathématique, c) maxima régionaux

5. **Modification des minima régionaux par gradient**

Pour modifier une image de telle sorte qu'il a un minima régionale seulement. Ici, vous pouvez utiliser la fonction "imimposemin" pour modifier l'image gradient d'amplitude de sorte que ses seuls minima régionaux se produisent au premier plan.

#### 6. Segmentation par Ligne de partage d'eau

À ce stade, on applique la LPE sur l'image gradient de l'étape (5), comme indiquée dans la figure (4.11.a). Les marqueurs et les contours superposés sur l'image d'origine sont indiqués sur la figure (4.11.b)

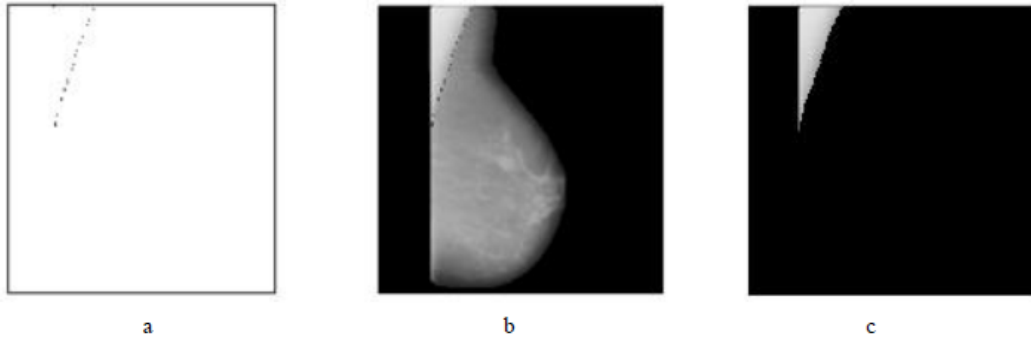


FIGURE 4.11 – Principales étapes de segmentation du muscle pectoral : a) LPE de gradient modifié, b) détermination du muscle pectoral par LPE, c) segmentation du muscle

La suppression du muscle pectoral se fait ainsi en multipliant l'image ( figure 4.9.a) par l'image binaire ( figure 4.12.b). La figure (4.12.c) montre le résultat obtenu après la suppression du muscle pectoral.

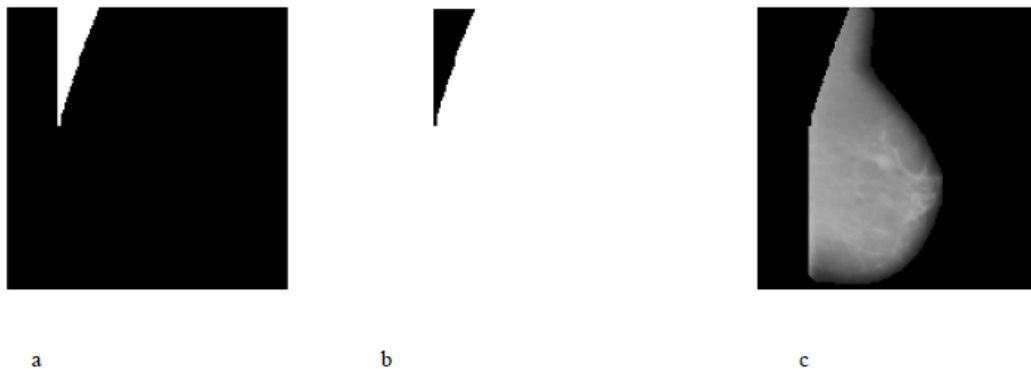


FIGURE 4.12 – Segmentation du muscle pectoral(a), b) binarisation du muscle pectoral, c) suppression du muscle pectoral

Les figures (4.13) et (4.14) montrent respectivement les résultats de segmentation du muscle pectoral par la méthode de Jayasree Chakraborty et al [4] et la méthode proposée . Une autre comparaison de la segmentation de l'image mdb199 par notre approche et les trois algorithmes de Ferrari et al [109], Kwok et al [114] et Jayasree Chakraborty et al [4] est présentée sur la figure (4.15).



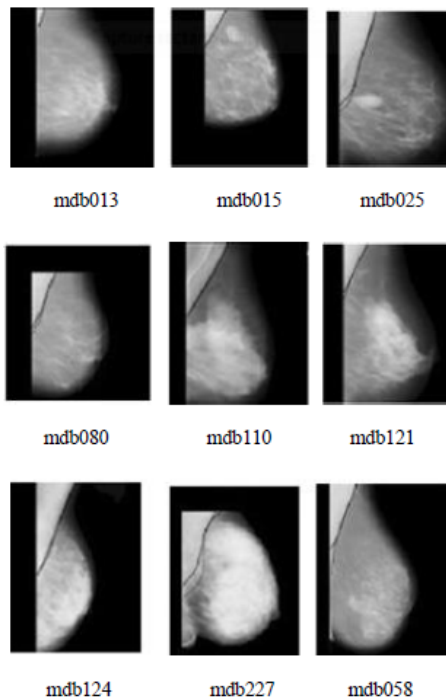


FIGURE 4.13 – Résultats de la détection du muscle pectoral par la méthode Jayasree Chakraborty [4]

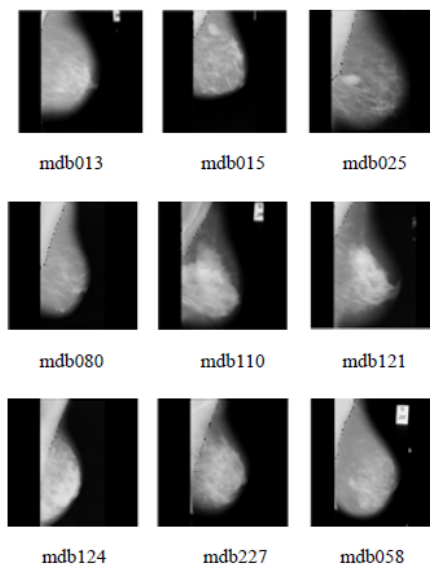


FIGURE 4.14 – Segmentation du muscle pectoral par la méthode proposée

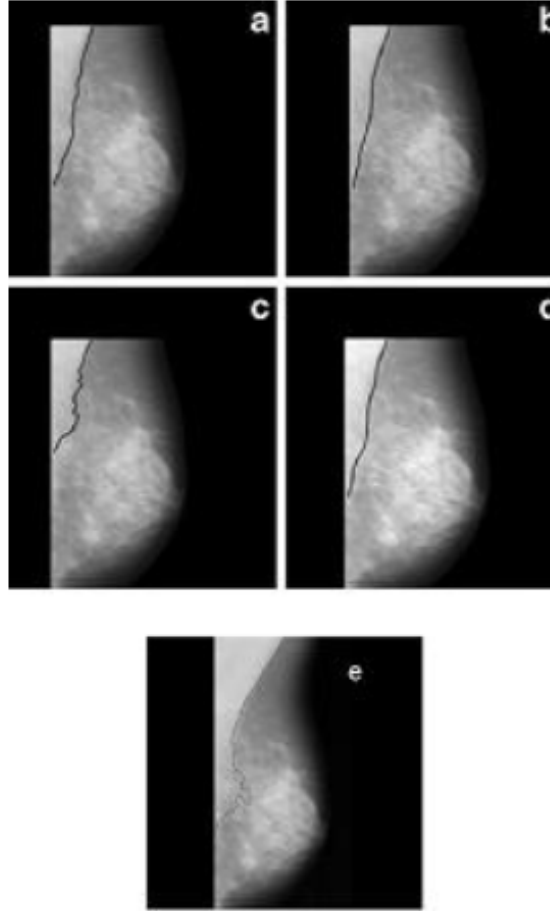


FIGURE 4.15 – Segmentation du muscle pectoral de l’image mdb199 : a) manuellement, b) par la méthode Jayasree Chakraborty, c) par la méthode de Ferrari, d) par la méthode de Kwok, et e) par la méthode proposée

### 1.3. Evaluation des performances :

Pour mesurer la performance de notre approche, nous utilisons les paramètres suivants :

- Faux positif (FP) est le pourcentage de pixels défini à l’intérieur de la limite du muscle pectoral obtenu mais en dehors du muscle pectoral de référence :

$$FP = \frac{|P \sqcup R| - |R|}{|R|} \times 100 \quad (4.3)$$

- Faux négatifs (FN) est le pourcentage de pixels défini à l’intérieur de la limite de muscle pectoral de référence mais en dehors de la limite de muscle pectoral obtenu normalisée par les pixels totaux de la région de muscle pectoral de référence.

$$FN = \frac{|P \sqcup R| - |P|}{|R|} \times 100 \quad (4.4)$$

- Vrais positifs (VP) est le pourcentage de pixels défini à l’intérieur de la limite du muscle pectoral obtenu et de référence :

$$VP = \frac{|P \cap R|}{|R|} \times 100; \quad (4.5)$$

— Vrais négatifs (VN) est le pourcentage de pixels défini à l'extérieur de la limite du muscle pectoral obtenu et de référence :

$$VN = \frac{|P \cup R|}{|R|} \times 100; \quad (4.6)$$

Où,

$P(x,y) \in$  la région obtenu du muscle pectoral

$R(x,y) \in$  la région référence du muscle pectoral

— Total mismatched (Tot) est obtenu à partir des FP et FN et peut être défini par :

$$Tot = \frac{2 * |P \sqcup R| - (|P| + |R|)}{|R|} \times 100\% \quad (4.7)$$

## 1.4. Résultats et discussions

L'algorithme proposé a été testé sur 80 images mammographiques numériques tirées de la base de données MIAS. Pour évaluer l'efficacité de cette méthode, nous avons utilisé les paramètres d'évaluation de la performance mentionnés précédemment. La méthode proposée comparée aux trois algorithmes de détection du muscle pectoral à savoir la méthode proposée par Jayasree Chakraborty, Ferrari et Kwok. Le tableau (4.1) représente les résultats obtenus :

Tableau 4.1 – Moyenne de FP, FN, et le (Tot) de la base mini-MIAS.

Mesure de la performance	Méthode Proposée	Chakraborty	Ferrari	Kwok
FP	2.98	1.84	4.64	1.42
FN	3.68	6.56	3.93	8.36
(Tot)	6.67	8.4	8.57	9.78
Nbre d'images avec (FP et FN) <5%	47	38	37	24
Nbre d'images avec (FP et FN) <10%	80	62	53	50

Les graphes des figures (4.16), (4.17), (4.18) et (4.19) présentent les résultats de l'évaluation pour des images de la base de données MIAS.

La figure (4.16) présente l'évaluation de la performance de la méthode proposée par les mesures VP et FP. D'après la figure, nous constatons que la mesure de la VP est variée pour les différentes images dans un intervalle de 0.90 à 0.99, et pour le FP est variée dans un intervalle de 0.0001 à 0.14.

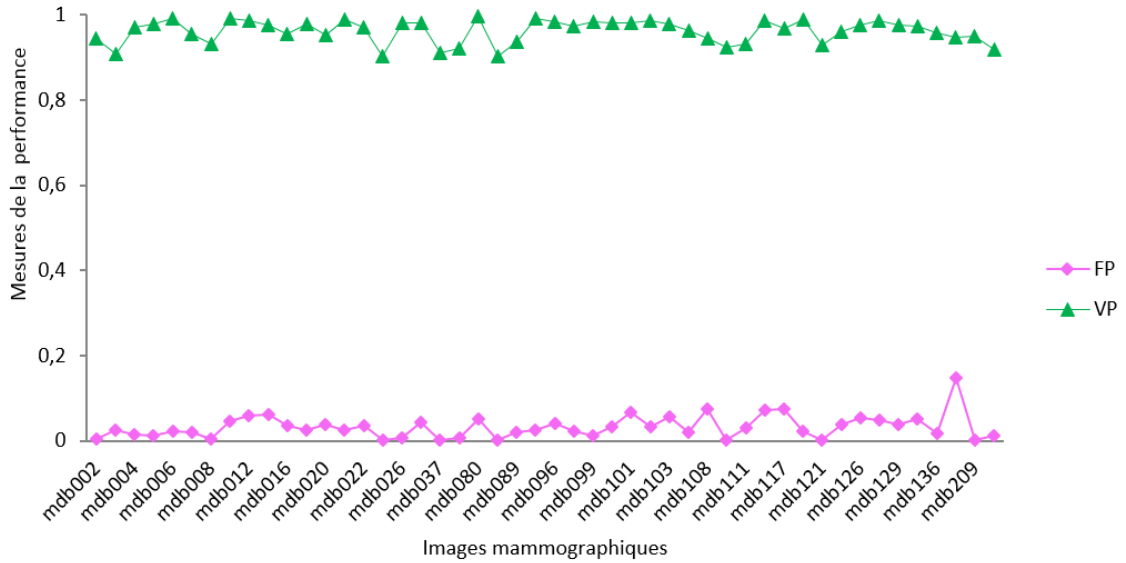


FIGURE 4.16 – Résultat de l'évaluation avec vrai positif et de faux positif

La figure (4.17) présente l'évaluation de la performance de la méthode proposée par la mesure de la sensibilité (Se) et la spécificité (Sp) (voir équation 3.60 et 3.61). D'après la figure, nous constatons que la mesure de la sensibilité (Se) qui reflète les performances d'un algorithme pour détecter les muscles pectoraux est variée pour les différentes images dans un intervalle de 0.90 à 0.99, et pour la spécificité (Sp) qui évalue la capacité de détecter les muscles non pectoraux est variée dans un intervalle de 0.87 à 0.90.

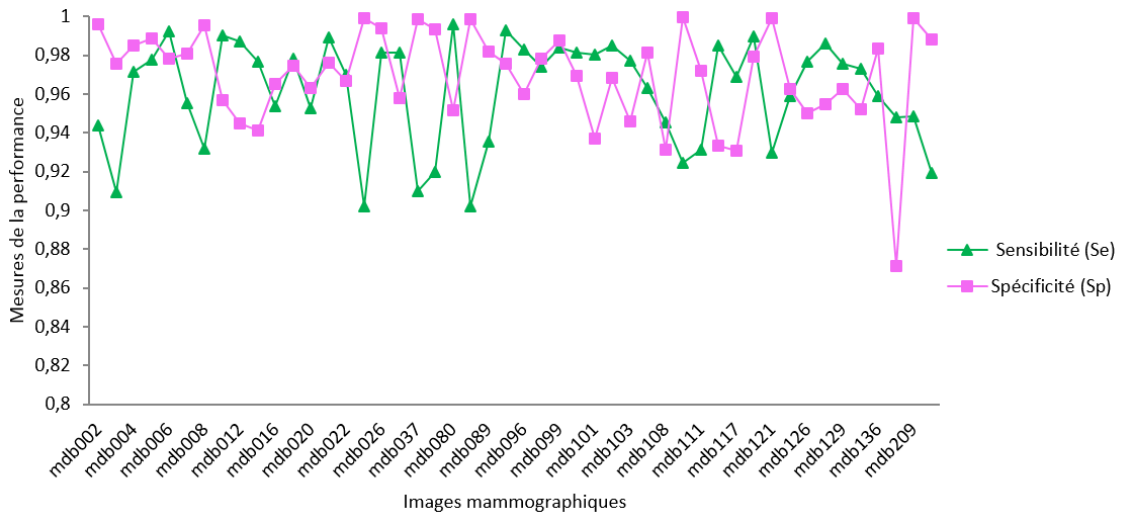


FIGURE 4.17 – Résultat de l'évaluation avec sensibilité et spécificité

La figure (4.18) présente l'évaluation de la performance de la méthode proposée par la mesure de l'exactitude (Ac) et la précision (Pr) (voir équation 3.64 et 3.65). D'après la fi-

gure, nous constatons que la mesure de l'exactitude qui présente la mesure le rapport entre le nombre total des pixels correctement classés (somme des vrais positifs et vrais négatifs) et le nombre des pixels dans l'image (somme des vrais positifs, vrais négatifs, faux positifs et des faux négatifs)est varie pour les différentes images dans un intervalle de 0.90 à 0.99. Et pour la précision, elle varie dans un intervalle de 0.86 à 0.99. La figure (4.19) présente la mesure de l'erreur (voir équation 3.66 et 3.67) qui est varie dans un intervalle de 0.01 à 0.05.

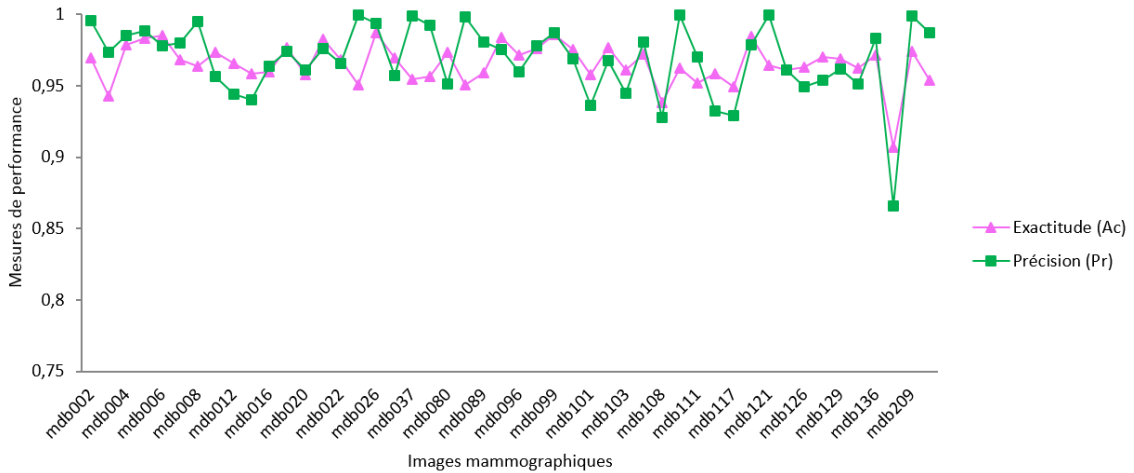


FIGURE 4.18 – Résultat de l'évaluation avec l'exactitude et la précision

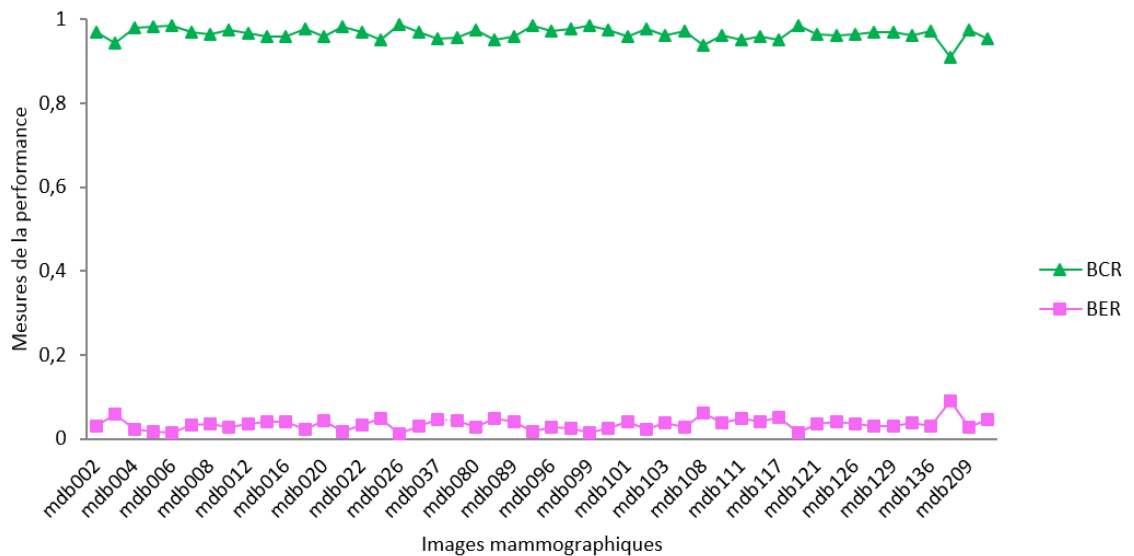


FIGURE 4.19 – Résultat de l'évaluation avec le taux d'équilibre de classification et le taux d'erreur

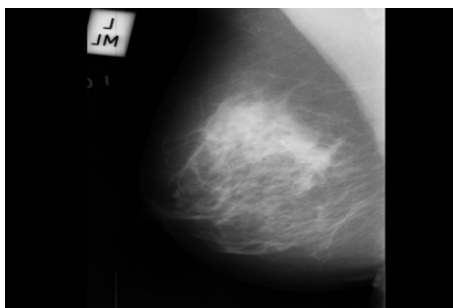
## 2. Détection des microcalcifications

Concernant la détection des microcalcifications basée sur l'analyse des images mammographiques, trois problèmes majeurs sont rencontrés à savoir : 1)- la taille des microcalcifications qui est très faible; ils sont visibles sous forme des petits points lumineux dans les images mammographiques, 2)- le fond inhomogène et 3)- le contraste qui est généralement faible.

Alors pour résoudre ces problèmes, nous proposons des techniques pour la détection des microcalcifications dans les images mammographiques fortement bruitées et peu contrastées. L'image doit être l'objet d'un prétraitement basé sur des méthodes adaptatives de rehaussement de contraste. Les images bruitées et rehaussées ainsi obtenues seront segmentées afin d'extraire les clusters des microcalcifications. Dans ce cadre, le filtre homomorphe rehausse les zones suspectes en nettoyant les encombrements liés à l'arrière plan de l'image. Le processus de segmentation intervient alors sur l'image rehaussée pour localiser les régions de microcalcifications suspectes en utilisant un seuillage des niveaux de gris d'intensité différente. Dans cette partie, nous avons testé plusieurs types d'ondelettes, pour valoriser leurs apports dans la détection des microcalcifications (Mcs). La transformée en ondelette, permet de représenter l'image d'une façon hiérarchique, ce qui facilite son interprétation à différentes échelles.

### 2.1. Prétraitement

Le prétraitement est une première étape qui consiste à améliorer les caractéristiques des images qui sont bruitées, de manière à faciliter la visualisation de la structure à détecter. La figure (4.20) présente une image contenant des microcalcifications qui apparaissent sous la forme de pixels dont le niveau de gris est très fort ce qui indique la présence d'une anomalie. Cette anomalie peut être un début de cancer. Notre but dans cette partie est la mise en valeur de ces microcalcifications par un processus de rehaussement.



---

FIGURE 4.20 – Image mammographique

#### 2.1.1. Extraction des régions d'intérêt (ROI)

La glande mammaire est une des régions de plus grande taille sur le cliché radiographique. Alors, pour réduire le temps de traitement nous proposons d'extraire les régions d'intérêt de l'image.

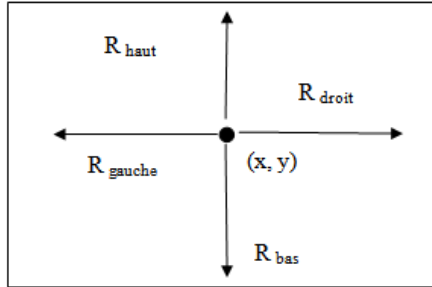


FIGURE 4.21 – Extraction d'échantillons en utilisant différentes "carrés" tailles de ROI

Le ROI des images mammographiques segmentées sur la figure ( 4.22) sont extraites par le découpage de la figure (4.21). Le lien<sup>1</sup> fournit des informations sur la nature et emplacement des microcalcifications. Le lien ci-dessus donne les zones qui contiennent des microcalcifications. Du centre, nous pouvons extraire des régions d'intérêt. Les images originales sont de taille 1024 \* 1024, les régions d'intérêt peuvent être soit 256\*256, 128\*128 ou 64\*64 en fonction du choix.

À partir de la figure (4.21) nous observons que les valeurs de décalage représentant la taille du ROI carré sont :  $R_{haut}$ ,  $R_{bas}$ ,  $R_{gauche}$  et  $R_{droite}$ . Les décalages utilisés pour extraire les ROIs de la taille 48 \* 48 pixels sont :  $R_{haut}=(x,y+24)$ ,  $R_{bas}=(x,y-23)$ ,  $R_{gauche}=(x-23,y)$  et  $R_{droite}=(x+24,y)$ . De même, pour extraire les ROIs de taille 256\*256 sont  $R_{haut}=(x,y+128)$ ,  $R_{bas}=(x,y-127)$ ,  $R_{gauche}=(x-127,y)$  et  $R_{droite}=(x+128,y)$ . La procédure appliquée pour l'extraction de ROI de la taille 256\*256 est défini par l'équation (4.8).

$$ROI = im(x - 127 : x + 128, y - 127 : y + 128) \quad (4.8)$$

avec

ROI : l'image après l'extraction de la région

im : l'image original de base MIAS

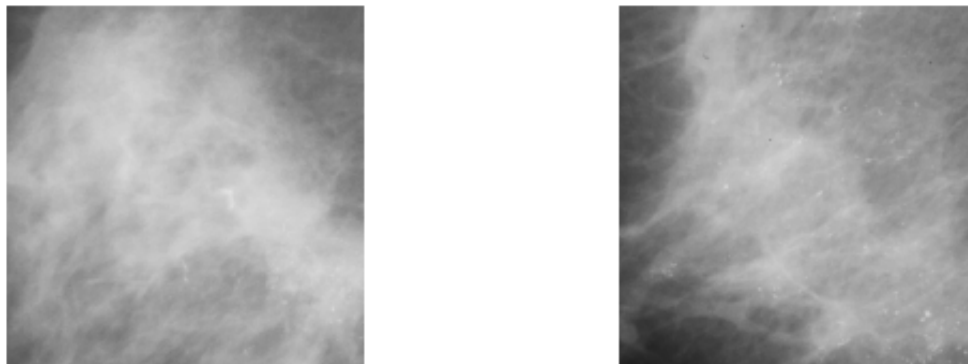


FIGURE 4.22 – Extraction du ROI 256 \* 256 de deux images différentes

1. [http : //peipa.essex.ac.uk/info/mias.html](http://peipa.essex.ac.uk/info/mias.html)

### 2.1.2. Rehaussement du contrast

On applique le rehaussement des images afin de faciliter l'interprétation visuelle et la compréhension des images.

L'histogramme de l'image est tout "ratatiné" : du coup, on n'y voit presque rien. Cela peut rendre l'image très sombre ou bien très grise ou encore trop claire ( voir figure 4.23(a, c)). Pour pallier ce phénomène, on va "dé-ratatiné" l'histogramme de notre image. On appelle cette opération "étirement d'histogramme", ou encore "normalisation des niveaux de gris / des couleurs". Le but de la manoeuvre, c'est de passer d'un histogramme tel que celui-ci (figure 4.23(b, d)).

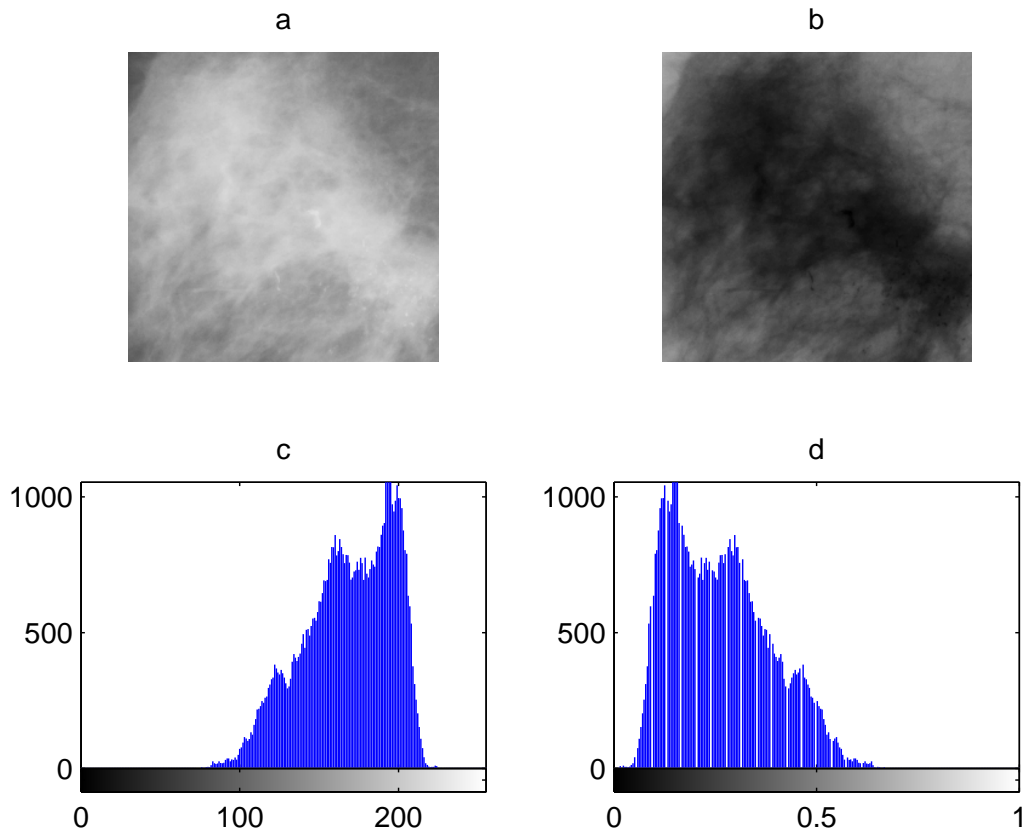


FIGURE 4.23 – Résultat de normalisation : a) Image originale, b) Normalisation, c) Histogramme de l'image originale, d) Histogramme de l'image normalisée

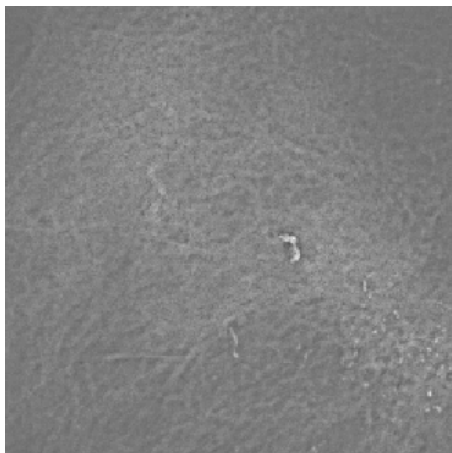
Nous constatons clairement l'effet de la normalisation, qui a rendu l'image bien contrastée.

### 2.1.3. Filtrage par un filtre haute fréquence

Le filtre passe-haut a tendance à retenir l'information à haute fréquence dans une image tout en réduisant les informations de basse fréquence (voir le chapitre 2 section(1.1.2)). Pour cela, nous avons utilisé un filtre homomorphique qui permet de garder les petits détails d'une image correspondant à des hautes fréquences et amortie les basses fréquences. Après le



filtrage de l'image, on voit clairement sur la figure (4.24) que l'image est toute grise excepté les Mcs.



---

FIGURE 4.24 – Image après filtre homomorphique

## 2.2. Détection des microcalcifications

### 2.2.1. Analyse par multirésolution des transformée en ondelettes

L'un des outils d'analyse multirésolution qui a été largement utilisé en le traitement d'images est l'analyse par ondelettes, proposée sous la forme de l'algorithme de pyramide de Mallat ; où une image peut être décomposée en sous-bandes de détails au niveau des résolutions différentes. La décomposition a été effectuée par filtrage des images à l'aide du filtre passe-bas (G) et passe-haut (H), suivi par un sous-échantillonnage d'un facteur de 2. Cette décomposition de l'ondelette est connue sous le nom de la transformée en ondelettes discrètes séparables (DWT) de deux dimensions (2D). Notre procédure de détection est la décomposition de l'image originale en sous-bandes à bas-bas Approximation (A), bas-haut vertical (V), haut-bas horizontal (H), et de haut-haut diagonale (D) comme illustré dans la figure (4.25). Dans l'ensemble du système la sous-bande A1 plus décomposé en quatre autres sous-bande A2, V2, D2 et H2. Puis la sous-bande A2 se décompose à nouveau. Trois niveaux de décomposition sont nécessaires parce que la plus faible bande de fréquence qui est générée est mise à zéro comme illustré à la figure (4.26), puisque les autres sous-bande contiennent les informations microcalcifications haute fréquence. Après l'étape de décomposition, on reconstruit l'image et la procédure de détection de microcalcifications peuvent commencer pour obtenu une image qui contient seules les informations à haute fréquence comme illustré à la figure (4.27).

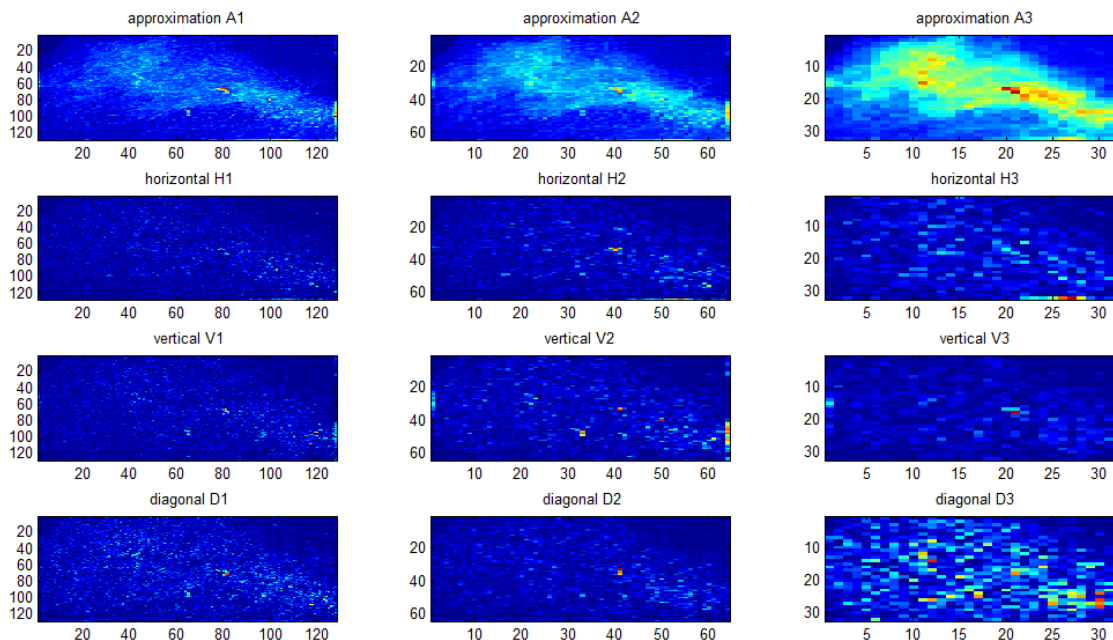


FIGURE 4.25 – Les approximations et les détails

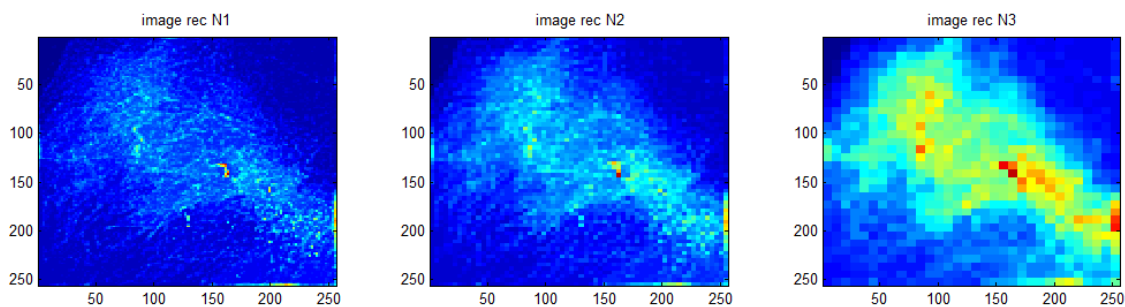


FIGURE 4.26 – Annulation des détails

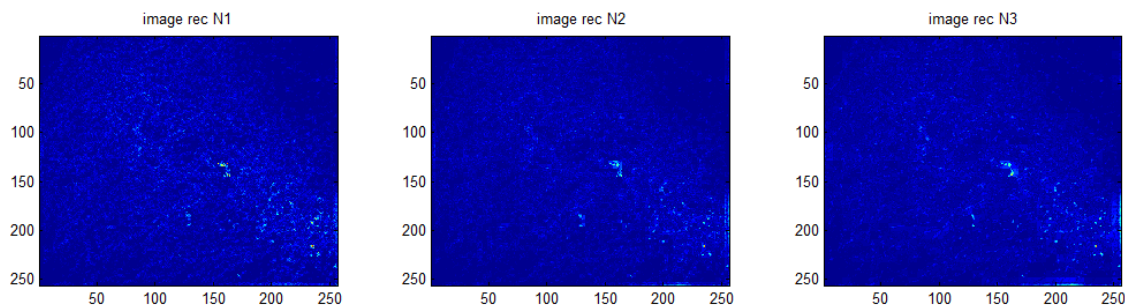


FIGURE 4.27 – Annulation des approximations

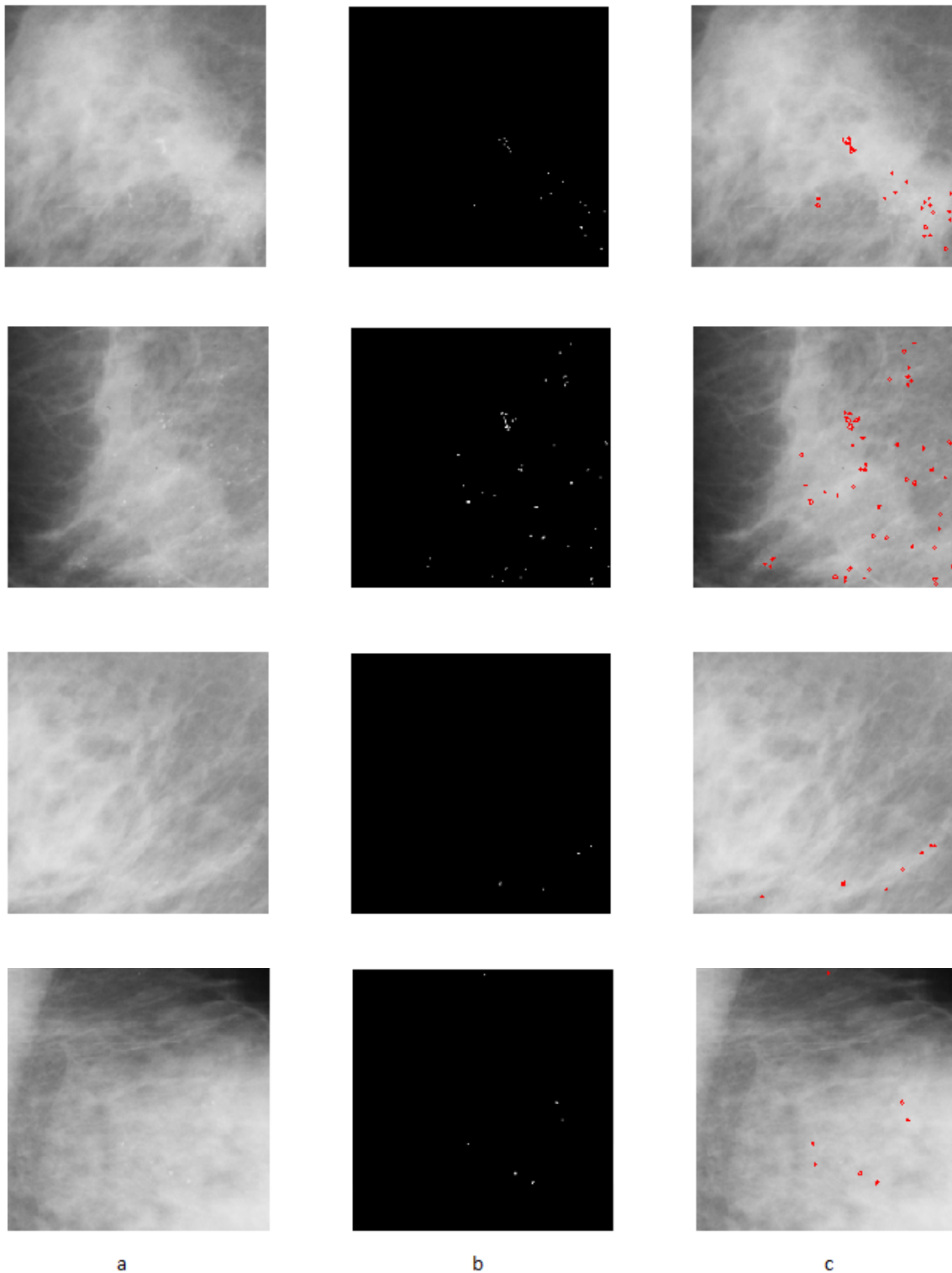


FIGURE 4.28 – Segmentation des microcalcifications : a) Image originale, b) image après le seuillage, c) contour coloré de l'image (b) superposé sur l'image originale (a)

La figure (4.28.a) représente l'image originale avant le prétraitement. La figure (4.28.b) donne l'image binaire après annulation des coefficients d'approximation au niveau 2 et la

figure (4.28.c) représente le contour coloré de l'image (voir figure 4.28.b) superposé sur l'image originale (voir figure 4.28.a).

### 3. Classification des images mammographiques

Pour mesurer l'exactitude d'un test médical, nous devons l'évaluer sur un ensemble de patients atteints ou non de la maladie pour confirmer les résultats obtenus. Quatre cas de figures peuvent se présenter selon le tableau (4.2)(ou matrice de confusion) :

Tableau 4.2 – Matrice de confusion

Réponse du classifieur	Réponse de l'expert	
	Positifs	Negatifs
Positifs	VP(Vrais Positifs)	FP(Faux Positifs)
Negatifs	FN(Faux Negatifs)	VN(Vrais Negatifs)

- VP : un groupe de personnes ont la maladie, et le test confirme la présence de la maladie.
- FN : un groupe de personnes ont la maladie, et le test ne le confirme pas la présence de la maladie.
- VN : un groupe de personnes n'ont pas la maladie, et le test confirme l'absence de la maladie.
- FP : un groupe de personnes n'ont pas la maladie, et le test confirme la présence de la maladie.

Pour évaluer notre approche proposée, nous l'avons tester à l'aide des trois classifieurs : les RN, le KPPV et le SVM. Deux bases de données ont été utilisées dans nos expérimentations de tailles respectives 64 et 140 images mammographiques de la base MIAS. Le tableau suivant (4.3) illustre comment les bases ont été utilisées dans nos expérimentations :

Tableau 4.3 – Base des images mammographiques

	Anormale	Normale	Apprentissage	Test
DB <sub>64</sub>	32	32	18	46
DB <sub>140</sub>	86	54	94	46

Nous avons utilisé l'exactitude de classification (AC) (voir l'équation 3.64) comme mesure de performance de nos tests.

#### 3.1. Classification des images normales et anormales avec les ondelettes

##### 3.1.1. Effet du prétraitement sur les résultats de la décision

Nos expérimentations ici porte sur l'évaluation de l'effet du prétraitement sur la bonne classification des images mammographiques. La figure (4.29) compare les résultats de classification avec l'algorithme de Kppv sans et avec prétraitement. La classification avec prétraitement

est bien améliorée par rapport au cas sans prétraitement. Les meilleurs taux de classification, environ 85%, sont obtenus pour  $k=1$  et 2.

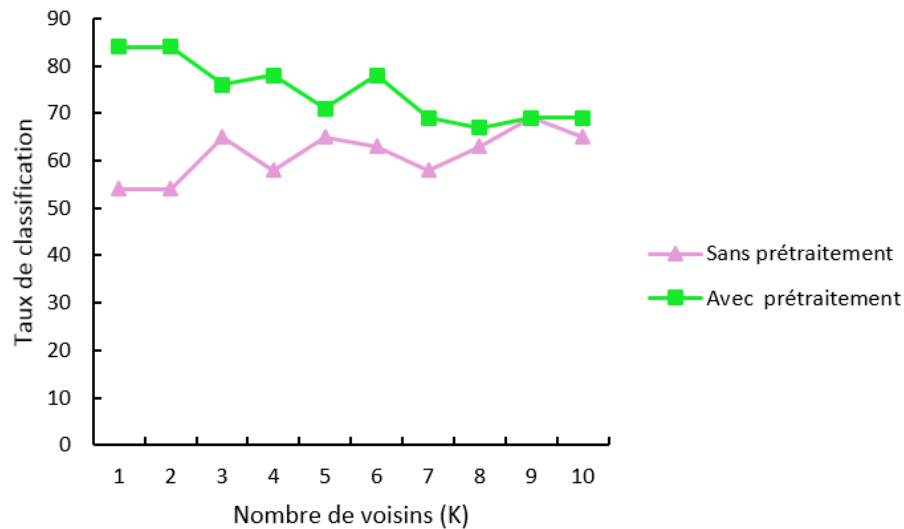


FIGURE 4.29 – Courbes de classification avec et sans prétraitement

### 3.1.2. Effet des niveaux de résolution sur les résultats de la décision

Nous avons testé les différents types d'ondelettes visant à incrémenter le pourcentage d'énergie qui correspond aux détails horizontaux, verticaux et diagonaux du troisième niveau de la transformée en ondelettes.

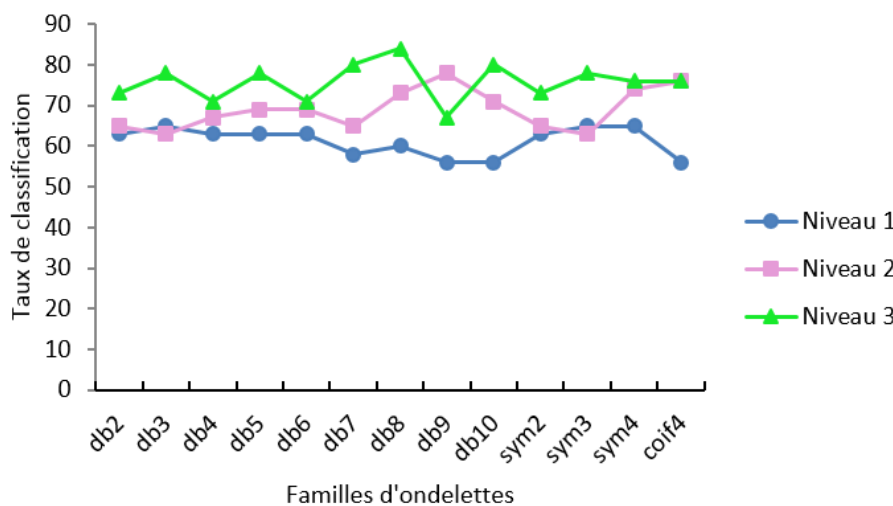


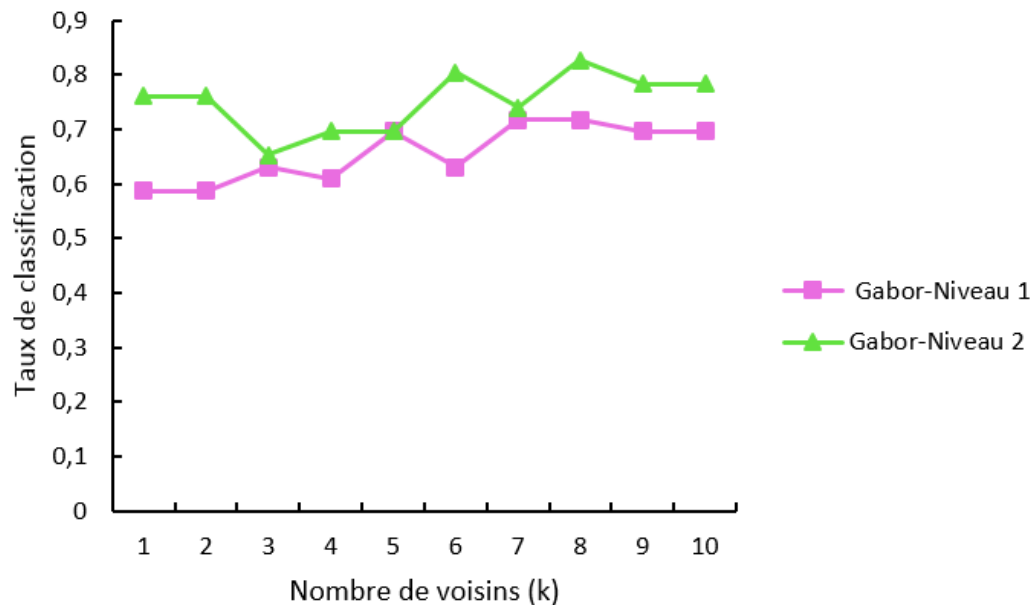
FIGURE 4.30 – Les résultats des expériences de l'application des différentes familles d'ondelettes sur les images mammographiques.

La figure (4.30) présente une comparaison entre les ondelettes orthogonale, afin de

décider de la famille appropriée à utiliser pour le traitement d'images mammographiques. L'amélioration de l'image mammographique est une solution clé en vue de réaliser la détection automatique des microcalcifications et de favoriser le pronostic précoce du cancer du sein. Le critère d'évaluation de la décision de l'ondelette est le degré de similitude entre les ondelettes et le profil d'image. Ceci a été évalué par le pourcentage d'énergie qui correspond aux coefficients de la transformée en ondelettes. Nos expériences ont montré que les ondelettes de Daubechies avec une longueur asymétrique 8 sont les plus appropriées, étant donné qu'elles accumulent plus d'énergie correspondant aux détails de la transformée en ondelettes. On peut en conclure que le taux de classification est obtenu par les caractéristiques extraites au troisième niveau de décomposition .

Trois expériences ont été classifiées à l'aide de kppv. Dans l'expérience 1, la classification est faite selon les caractéristiques Gabor. Dans l'expérience 2, la classification est faite selon les caractéristiques Tamura. Et dans l'expérience 3, nous avons extraits les caractéristiques Gist pour effectuer la classification.

Comme nous l'avons introduit dans (le chapitre 3), le classifieur k plus proche voisins dépend du choix du paramètre k. Dans cette section, nous allons étudier le comportement de la classification par k-ppv pour différentes valeurs de k et par le niveau de décomposition d'ondelette. Nous pouvons alors comparer les performances de classification en fonction du type d'information extrait par différents descripteurs et évaluer l'influence du nombre k dans l'algorithme Kppv.



---

FIGURE 4.31 – Le taux de classification pour gabor

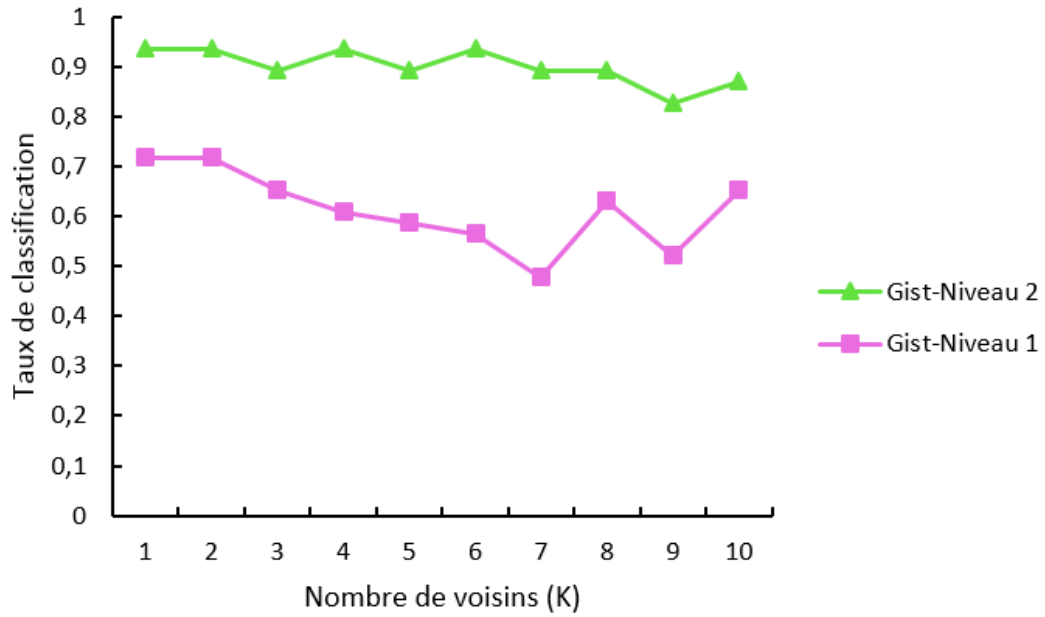


FIGURE 4.32 – Le taux de classification pour gist

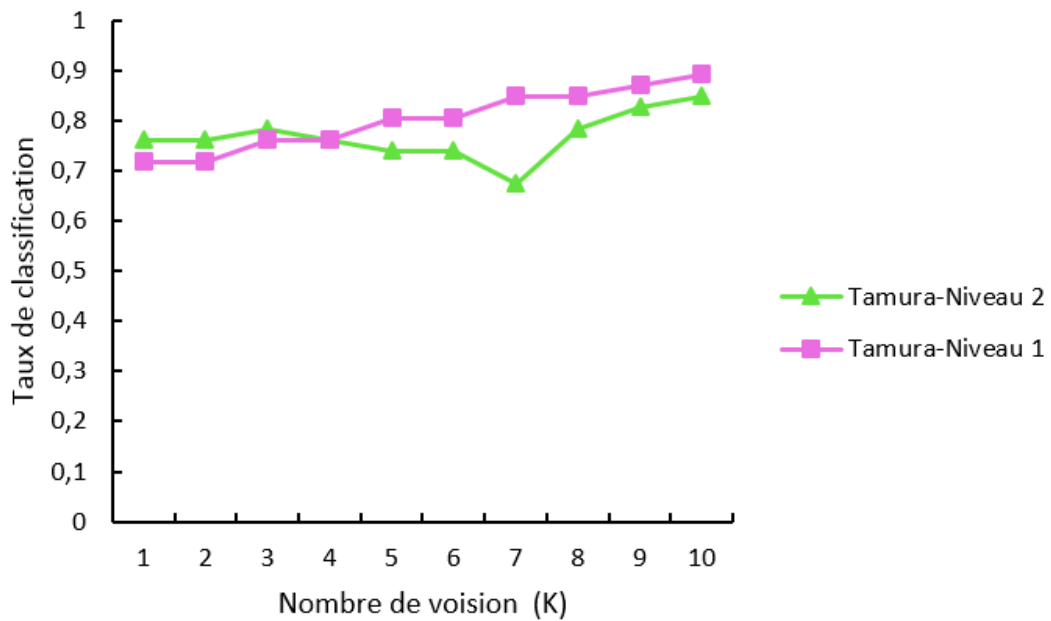


FIGURE 4.33 – Le taux de classification pour tamura

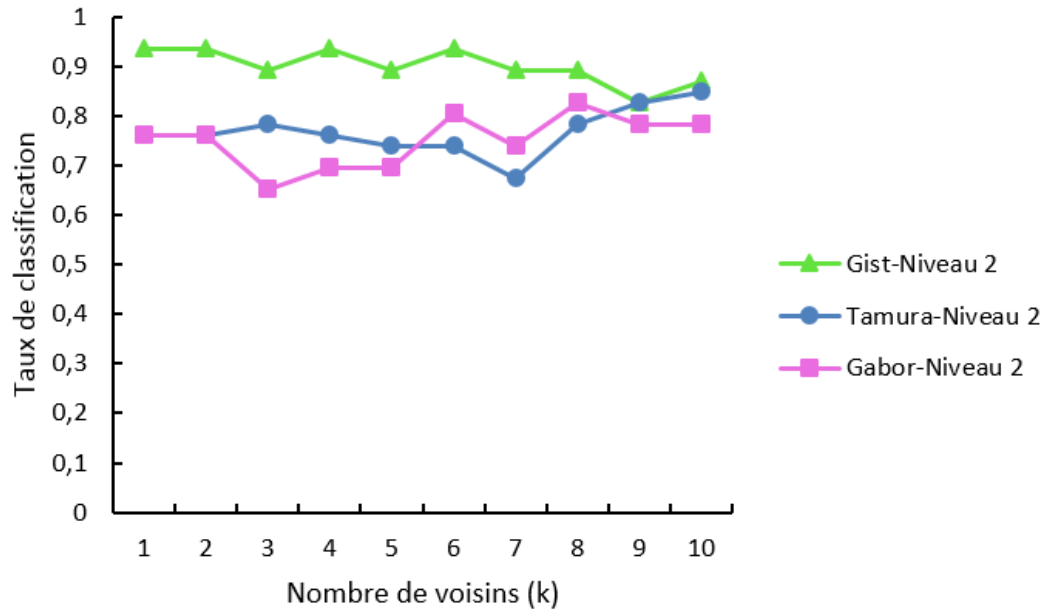


FIGURE 4.34 – Le taux de classification pour les trois descripteurs

Les figures (4.31), (4.32), (4.33) tracent le taux de classification du kppv pour différentes valeurs de  $k$  pour les descripteurs de Gist, Gabor et Tamura. Ainsi, nous remarquons que le meilleur taux de classification est obtenu pour  $k = 1$ . Nous pouvons facilement déduire que l'extraction des descripteurs à partir d'une image qui a subi une transformation par ondelette au niveau 2 donne des performances nettement meilleures que si les descripteurs sont calculés directement sur l'image original ou sur une image qui a subi une transformation par ondelette au niveau 1. Et d'après la figure (4.34), nous constatons que les descripteurs de Gist donnent un meilleur taux de classification par rapport aux autres descripteurs pour différentes valeurs de  $k$  de kppv.

Nous avons résumé les résultats que nous avons obtenus après la phase de la classification dans le tableau (4.4) ci-dessous qui donne le taux de classification pour chaque base et pour chaque descripteur utilisé.

Tableau 4.4 – Evaluation des résultats avec différentes bases des images

Classifieurs Descripteurs	140 images		64 images	
	KPPV	RN	KPPV	RN
Gist	93	95	71.73	74
Dwt	84.78	86	67.39	63
Gabor	76	71	69.56	68
Tamura	76	89.78	43.47	74

À partir ce tableau (4.4), nous constatons que les descripteurs Gist donne de meilleurs taux de classification pour les deux bases, suivis de descripteurs DWT pour la 1<sup>ère</sup> base,



environ 85% en moyenne. Les descripteurs Tamura donnent un taux acceptable en utilisant les réseaux de neurones

### 3.1.3. Effet des classifieurs sur les résultats de la décision

Le tableau (4.5) résume les résultats des tests de classification effectués sur la 1<sup>ère</sup> base. Cinq descripteurs et leurs combinaisons ont été testés sur trois classificateurs (KPPV, RN et SVM).

Tableau 4.5 – Evaluation des résultats avec différents classifieurs

Classifieurs \ Descripteurs	base1 :140 images		
	KPPV	RN	SVM
Gist	93	95	86
Dwt	84.78	86	84
Gabor	76	71	71
Tamura	76	89.78	80
Matrice de co-occurrence	84	60	95
Tamura et Dwt	91	89	86
Gabor et Tamura	78	82	86
Tamura, Dwt et Gabor	78	84	89

À partir de ces résultats, nous constatons que les réseaux de neurones donne un meilleur taux de classification en utilisant les descripteurs de Gist, DWT et Tamura. Pour les matrices de co-occurrence, la classification par SVM performe les deux autres classifieurs. La combinaison (Tamura, DWT) donne de bons résultats avec le classificateur KPPV. Pour le reste des combinaisons, c'est le classificateur SVM qui performe encore.

## 3.2. Classification des images normales et anormales avec les contours non sous échantillonnées

### 3.2.1. Approche hybride combinant la transformation de Contourlet non sous échantillonné et le filtre homomorphique

La figure(4.35) présente une comparaison de l'amélioration des images mammographiques basée sur trois algorithmes différents : filtre homomorphique, masquage flou et la méthode proposée où nous avons combiné la transformation de contourlet non échantillonné avec le filtre homomorphique.

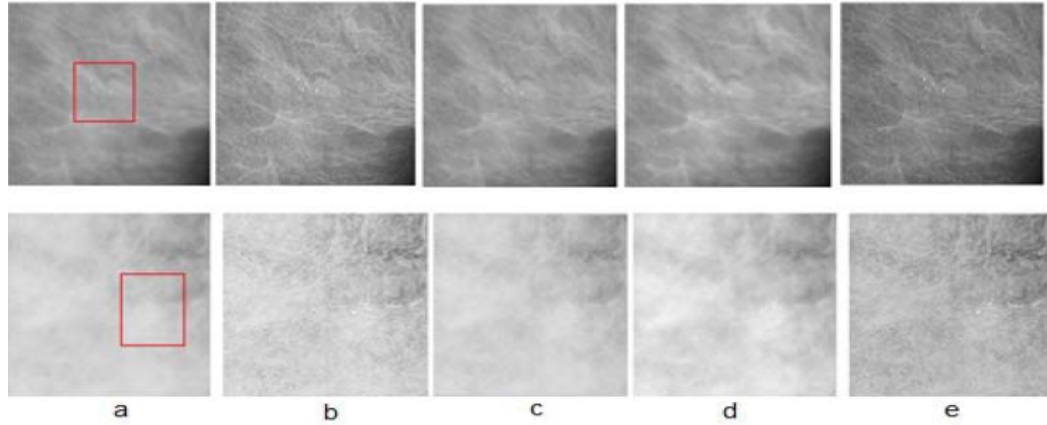


FIGURE 4.35 – Comparaison de différentes techniques d’amélioration : a) ROI de l’image originale, b) le NSCT, c) le filtrage homomorphique, d) le masquage flou, e) notre approche proposée

Pour évaluer notre méthode proposée, nous avons effectué plusieurs expérimentations où nous avons comparé trois méthodes avec celle proposée à savoir le NSCT, le filtrage homomorphique et le masque flou. Les Résultats expérimentaux montrent que l’approche proposée réduit les bruits dans les images mammographiques.

### 3.2.2. Techniques d’évaluation d’amélioration du contraste pour mammographique

Pour évaluer la qualité des images, nous avons utilisé trois mesures quantitatives, proposées par S. Singh et K. Bovis [48], la mesure du contraste basée sur la moyenne (Distribution Separation Measure (DSM)), la mesure du contraste basée sur la déviation standard (TBCS) et la mesure du contraste basée sur la de entropie (TBCE).

— **Evaluation des résultats avec (Distribution Separation Measure (DSM))**  
Le DSM défini par l’équation (4.9).

$$DSM = (|\mu_T^E - \mu_B^E|) - (|\mu_T^O - \mu_B^O|) \quad (4.9)$$

Avec,  $\mu_T^O, \mu_T^E$  sont respectivement, la moyenne de la région cible avant et après l’amélioration,  $\mu_B^O, \mu_B^E$ , sont respectivement, la moyenne de la région de l’arrière-plan avant et après l’amélioration.

— **Evaluation des résultats avec (Target-to-Background Contrast Enhancement Measure Based on Standard Deviation (TBCS))**

Un objectif d’une amélioration du contraste est de maximiser la différence entre le fond et la cible du niveau moyen de gris et veiller à ce que l’homogénéité de la masse est accrue aidant la visualisation de ses limites et l’emplacement. En utilisant le rapport de la norme déviation des échelles de gris à l’intérieur de la cible avant et après l’amélioration, nous

pouvons quantifier cette amélioration en utilisant la TBCS donnée dans l'équation(4.10) :

$$TBCS = \frac{\frac{\mu_T^E}{\mu_B^E} - \frac{\mu_T^O}{\mu_B^O}}{\frac{\sigma_T^E}{\sigma_T^O}} \quad (4.10)$$

Avec,  $\sigma_T^O$ ,  $\sigma_T^E$  sont l'écart type des niveaux de gris comprenant la cible avant et après l'enhancement.

— **Evaluation des résultats avec (Target-to-Background Contrast Enhancement Measure Based on Entropy (TBCE))**

Cette mesure est une extension de la TBCS. La TBCE est basée sur l'entropie des régions plutôt que sur les écarts-types et est défini par l'équation (4.11).

$$TBCE = \frac{\frac{\mu_T^E}{\mu_B^E} - \frac{\mu_T^O}{\mu_B^O}}{\frac{\mathcal{L}_T^E}{\mathcal{L}_T^O}} \quad (4.11)$$

Avec,  $\mathcal{L}_T^O$ ,  $\mathcal{L}_T^E$ , sont l'entropie de la cible dans les images originales et améliorées.

### 3.2.3. Résultats expérimentaux

Les figures (4.36), (4.37) et (4.38) représentent respectivement les courbes des différentes mesures d'évaluation des DSM, TBCS et TBCE testées pour les quatres approches d'amélioration.

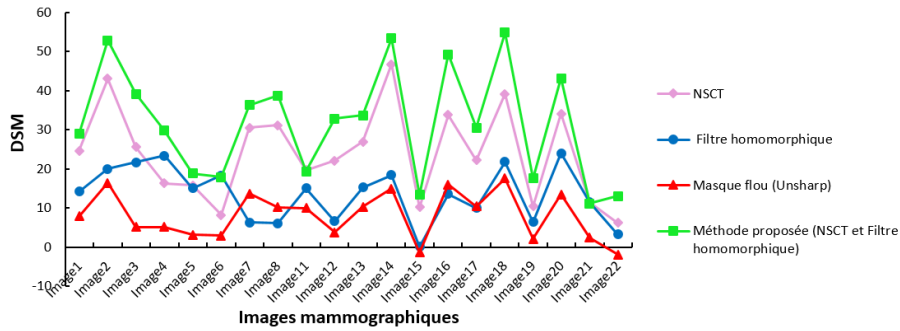


FIGURE 4.36 – Traçage de DSM

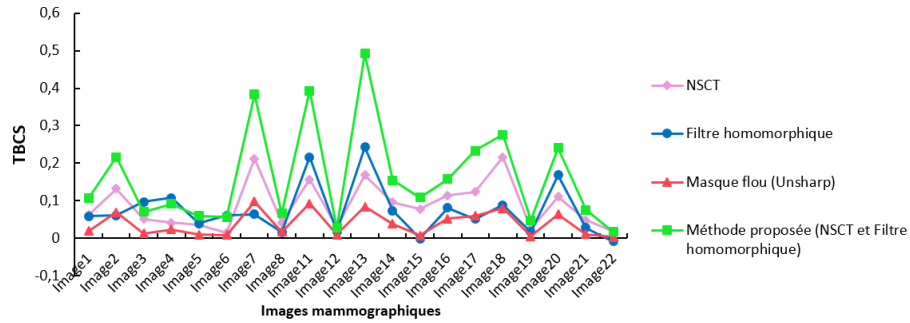


FIGURE 4.37 – Traçage de TBCS

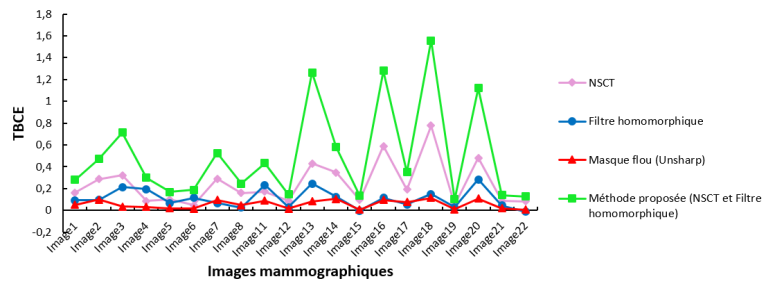


FIGURE 4.38 – Traçage de TBCE

À partir de ces figures, nous constatons que la méthode proposée, qui a pour principe hybrider la méthode NSCT et filtre homomorphique, donne des bons résultats comparée aux autres méthodes testées.

### 3.3. La phase de classification

Dans cette phase, nous avons utilisé le classificateur KPPV pour classifier les images mammographiques. La classification effectuée ici est de type binaire; anormal au cas de présence des microcalcifications ou normal dans le cas échéant. Des expériences ont été réalisées avec plusieurs valeurs de k pour le classificateur de kppv et plusieurs fréquences de coupure D0 nécessaire pour le filtre homomorphique.

La performance des approches proposées pour la classification des images normales et des images anormales est mesurée par la mesure d'exactitude.

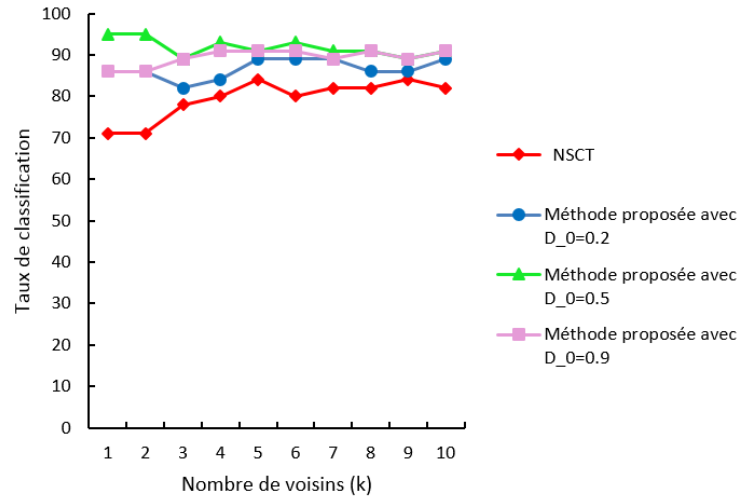


FIGURE 4.39 – Taux de classification de la méthode NSCT et la méthode proposée.

La figure (4.39) compare et évalue notre approche proposée avec la méthode NSCT en terme de taux de classification. Plusieurs valeurs du paramètres  $k$  du classificateur de kppv ont été testées ( $k=1, \dots, 10$ ). Aussi, nous avons testé trois fréquence  $D_0$  du filtre homomorphique pour notre approche,  $D_0=0.2$ ,  $D_0=0.5$  et  $D_0=0.9$ . D'après cette figure, nous constatons que les meilleurs taux sont obtenus toujours pour  $k=1$  et  $2$  et pour une fréquence de coupure  $D_0=0.5$ .

## Conclusion

Dans ce chapitre d'expérimentations, nous avons montré l'efficacité de la technique des lignes de partage d'eau et l'utilisation de la morphologie mathématique pour supprimer les étiquettes et le muscle pectoral du sein. Cette technique est la meilleure car elle assure l'identification et la formation de contour fermé de l'objet ; ce qui a facilité la détection des lésions mammaires ainsi l'opacité dans les mammographies numérisées. Ensuite, nous avons testé cinq méthodes d'extraction de caractéristiques où nos expériences ont montré que les descripteurs de Gist donnent un meilleur taux de classification en utilisant les réseaux de neurones et les descripteurs de la matrice co-occurrence avec le classificateur SVM donnent aussi le taux de classification de 95%. Et, pour finaliser ce travail de thèse avec notre dernière contribution qui consiste en l'amélioration du contraste des images mammographiques. Pour cela, nous avons proposés dans cette thèse une approche hybride des contourlets non sous échantillonné et le filtre homomorphique. Les résultats expérimentaux obtenus sont encourageants et montrent l'efficacité de notre approche comparée aux autres méthodes testées individuellement.

# Conclusion et perspectives

Malgré l'évolution des nouvelles technologies pour améliorer la qualité des images mammographiques numériques, l'interprétation et l'analyse de ces images restent une tâche difficile même pour un radiologue étant donné que ces images sont fortement texturées et/ou bruitées. D'où le besoin de développer des systèmes d'analyse et d'interprétation automatique des images mammographiques numériques comme outil d'aide à la décision augmentant la fiabilité de l'interprétation par un expert.

Les travaux présentés dans cette thèse visent le développement des techniques assistées par ordinateur pour la détection des microcalcifications et la classification des régions d'un mammogramme en régions positives et d'autres négatives par l'analyse des textures des tissus.

La méthode de prétraitement, proposée consiste à augmenter le contraste des microcalcifications afin de rendre leur détection plus performante et plus efficace.

Cette thèse a nécessité dans un premier temps une étude de l'anatomie du sein et le processus d'acquisition des mammogrammes numériques d'interprétation utilisés par les radiologues. Ces derniers, vu leur expérience arrivent visuellement à localiser grossièrement les zones suspectes. Une segmentation automatique de ces zones permet de bien localiser les régions à analyser tout en limitant le temps de traitement et d'augmenter les performances du processus d'aide à la décision (minimiser les faux positifs). Ceci nous a incité à faire une recherche bibliographique très poussée sur les différentes méthodes de segmentation utilisées. En plus, une étude comparative de ces techniques a été à la base du choix des méthodes d'analyse statistique qui s'avèrent bien adaptées aux images mammographiques caractérisées par une texture aléatoire.

Pour localiser les régions d'intérêt, nous avons passé par plusieurs étapes préliminaires : la séparation du fond de la région du sein par la technique de seuillage et la morphologie mathématique ; ce qui a permis la détection du contour extérieur du sein ; l'élimination du muscle pectoral par la technique des lignes de partage d'eau pour minimiser les faux positifs et localiser les zones de forte densité. Ces étapes ont permis de mieux cerner la zone d'analyse tout en minimisant le temps de calcul.

Pour aboutir au but de recherche de microcalcifications, nous avons présenté une analyse plus fine par des méthodes de multirésolutions, qui permettent d'extraire les microcalcification en différents niveaux de résolution et par orientation. Les résultats expérimentaux ont montré que l'hybridation de normalisation et le filtre homomorphique est plus efficace pour éliminer les textures qui sont présentes dans les images mammographiques. Les zones d'intérêt sont finement analysées à l'aide d'un ensemble de descripteurs : 512 attributs extraits à partir de gist, 3 paramètres de gabor, 4 attributs de tamura, 8 paramètres extraits de la matrice de cooccurrence (Harralick) et 6 paramètres de dwt au niveau 2, qui caractérisent

de façon robuste les textures d'un mammogramme.

Une étude de l'influence de certains paramètres sur la détection des microcalcifications comme : la taille de la fenêtre d'analyse, l'orientation, les niveaux de résolution et l'étape du prétraitement des textures ont été effectuées avant tout type d'analyse. Cela a été nécessaire et nous avons aidé à retenir les combinaisons de paramètres qui donnent les meilleures performances. Toutes ces expérimentations ont été conduites à partir de la base MIAS des images mammographiques.

À ce stade, nous avons conclu que les descripteurs de Gist et les contourlets non-sous-échantillonnées donnent de meilleurs performances. À base de ces résultats encourageants, nous pensons qu'une extension de notre approche pour caractériser les pathologies en maligne et bénigne serait d'avantage intéressante.

# Bibliographie

- [1] H. Albert. *Épissage alternatif des phosphatases cdc25 dans le cancer du sein : expression des différents variants dans des modèles cellulaires et tissulaires de cancer mammaire et régulation en conditions de stress génotoxique. Thèse université paul verlaine-metz ufr sciences fondamentales et appliquées*, 26 octobre 2011.
- [2] J. Debayle. *Traitement d'image à voisinages adaptatifs généraux. Traitement du signal et de l'image. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2005. Français.*
- [3] M.N. Do and M. Vetterli. The contourlet transform : An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, pages 2091–2106, Sept 2002.
- [4] J. Chakraborty, S. Mukhopadhyay, V. Singla, N. Khandelwal, and P. Bhattacharyya. Automatic detection of pectoral muscle using average gradient and shape based feature. *Digit Imaging*, 2012.
- [5] Registre des cancers du service d'épidémiologie de l'institut national d'oncologie sidi mohamed ben abdellah, de rabat. 2002-2007.
- [6] Registre des cancers de la région du grand casablanca. 2005.
- [7] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, and F. Bray. Cancer incidence, mortality and prevalence worldwide. *International Agency for Research on Cancer IARC*, 69, 2004.
- [8] Registre des cancers. centre hospitalier d'oncologie d'oujda. 2007.
- [9] Registre des cancers. centre hospitalier d'oncologie de marrakech. 2007.
- [10] Guide de détection précoce des cancers du sein et du col de l'utérus edition. 2011.
- [11] I. Cheikhrouhou. Description et classification des masses mammaires pour le diagnostic du cancer du sein. *Thèse. Université D'EVRY VAL D'ESSONNE*, Juin 2012.
- [12] R. Touami. Segmentation des images médicales par ondelettes. *Mémoire de Magister. Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf*, 2011.
- [13] M. Sègbédji Rethice Théophile Junior Goubalan. Contribution à l'analyse d'images médicales pour la reconnaissance du cancer du sein. *thèse. École doctorale Sciences et technologies de l'information et de la communication*, 9 décembre 2016.
- [14] M. Teboul. Advantages of ductal echography (de) over conventional breast investigation in the diagnosis of breast malignancies. *Med Ultrason*, 12(1) :32–42, 2010.
- [15] A. Clément. Imagerie par résonances magnétiques irm : Topographie des articulations du coude, du capre, du grasset et du jarret de chiens sains. *Thèse de Doctorat, Ecole Vétérinaire De Nantes*, 10 Décembre 2009.



- [16] P. Kestener. Analyse multifractale 2d et 3d à l'aide de la transformation en ondelettes : application en mammographie et en turbulence développée. *PHD, l'université bordeaux I, Ecole Doctorale de Sciences Physiques et de l'Ingénieur*, Novembre 2003.
- [17] P. Hordé. Tumeur-définition. *des Femmes Santé*, Avril 2016.
- [18] P. Hordé. *Institut national du cancer*.
- [19] W.P. Evans. Breast masses appropriate evaluation. *The Radiologic Clinics of North America, Breast Imaging*, 33(6) :1085–1108, Sep 1995.
- [20] S.A. Feig and M.J. Yaffe. Digital mammography, computer-aided diagnosis and telemammography. *The Radiologic Clinics of North America, Breast Imaging*, 33(6) :1205–1230, sep 1995.
- [21] B.S. Monsees. Evaluation of breast microcalcifications. *The Radiologic Clinics of North America, Breast Imaging*, 33(6) :1109–1121, sep 1995.
- [22] M. Lanyi, J. Zander, and J. Baltzer. Morphologic analysis of microcalcifications. early breast cancer. *Springer-Verlag, Berlin*, 1985.
- [23] C.J. D'Orsi, L. W. Bassett, W. A. Berg, S. A. Feig, V. P. Jackson, M. N. Linver D. B Kopans, E. B. Mendelson, L. J. Moss, and E. A. Sickles. (breast imaging reporting and data system). troisième édition française réalisée par sfr (société française de radiologie). *American college of radiology*, 2003.
- [24] Comprendre le cancer du sein. *Institut national du cancer*, janvier 2007.
- [25] Registre des cancers de la région du grand casablanca. 2004.
- [26] T.J. Key, P.K. Verkasalo, and E. Banks. Epidemiology of breast cancer. *Lancet Oncology*, 2(3) :133–40, 2001.
- [27] MJ. Glade. Food, nutrition, and the prevention of cancer : a global perspective. *American Institute for Cancer Research/World Cancer Research Fund, American Institute for Cancer Research*, 15(6) :523–526, 1997.
- [28] F. Clavel-Chapelon and M. Gerber. Eproductive factors and breast cancer risk. do they differ according to age at diagnosis breast cancer res treat. *Lancet*, 72(2) :107–115, 2002.
- [29] D. Ford, D.F. Easton, and M. Stratton. Genetic heterogeneity and penetrance analysis of the brca1 and brca2 genes in breast cancer families. the breast cancer linkage consortium. *American Journal of Human Genetics - Cell Press*, 62(3) :676–689, 1998.
- [30] J.L.Berliner, A.M.Fay, SA. Cummings, B. Burnett, and T. Tillmanns. Risk assessment and genetic counseling for hereditary breast and ovarian cancer. *genetic counseling*, 22(2) :155–163, 2013.
- [31] A.W.Leung, J.Mak, and P.S.Cheung. Evidence for a programming effect of early menarche on the rise of breast cancer incidence in hong kong. *Cancer Epidemiology, Detection and Prevention*, 32(2) :156–161, 2008.
- [32] S.E.Singletary. Rating the risk factors for breast cancer. *Annals of Surgery*, 237(4) :474–82, 2003.
- [33] P.Irigaray, J.A.Newby, R.Clapp, L. Hardell, V. Howard, L. Montagnier, S. Epstein, and D. Belpomme. Lifestyle-related factors and environmental agents causing cancer : an overview. *Biomedicine and Pharmacotherapy*, 61(10) :640–658, 2007.

- [34] H.K.Seitz and P.Becker. Alcohol metabolism and cancer risk. *Alcohol Research and Health*, 30,(1) :38–41, 44–7, 2007.
- [35] JE. Rossouw, GL. Anderson, RL. Prentice, AZ. LaCroix, C. Kooperberg, and ML. Stefanick. Risks and benefits of estrogen plus progestin in healthy postmenopausal women : principal results from the women’s health initiative randomized controlled trial. *JAMA*, 288(3) :321–333, 2002.
- [36] GL. Anderson, M. Limacher, AR. Assaf, T. Bassford, SA. Beresford, and H. Black et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy : the women’s health initiative randomized controlled trial. *JAMA*, 291(14) :1701–1712, 2004.
- [37] Breast cancer and breastfeeding : collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet*, 360(9328) :187–195, 2002.
- [38] T.J. Key and MC. Pike. The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. *European Journal Cancer Clinical Oncology*, 24(1) :29–43, 1988.
- [39] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Geer, B. ter Haar Romeny, J.B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3) :355–368, 1987.
- [40] R. Sivaramakrishna, N.A. Obuchowski, W.A. Chilcote, G. Cardenosa, and K.A. Powell. Comparing the performance of mammographic enhancement algorithms : a preference study. *American Journal of Roentgenology*, pages 45–51, 2000.
- [41] W.M. Morrow, R.B. Paranjape, R.M. Rangayyan, and J.E.L Desautels. Region-based contrast enhancement of mammograms. *IEEE Transactions on Medical Imaging*, 11(3) :392–406, 1992.
- [42] T. L. Ji, M. K. Sundareshan, and H. Roehrig. Adaptive image contrast enhancement based on human visual properties. *IEEE Transactions on Medical Imaging*, 13(4) :573–586, 1994.
- [43] A. Laine, J. Fan, and W.H. Yan. Wavelets for contrast enhancement of digital mammography. *IEEE Engineering in Medicine and Biology Magazine*, 14(5) :536–550, 1995.
- [44] R. S. Pfisterer and F. Aghdasi. Hexagonal wavelets for the detection of masses in digitized mammograms. *SPIE*, 3813 :966–977, 1999.
- [45] A. Hassanien and J. Ali. Digital mammogram segmentation algorithm using pulse coupled neural networks. *International Conference on Image and Graphics*, pages 92–95, 2004.
- [46] N. Hamdi, K. Auhmani, and M. Hassani. Design of a high-accuracy classifier based on fisher discriminant analysis : Application to computer-aided diagnosis of microcalcifications. *Computational Sciences and Its Applications, 2008. ICCSA '08, IEEE*, pages 267–273, July 2008.
- [47] H. P. Chan, C. J. Vyborny, H. MacMahon, C. E. Metz, K. Doi, and E.A. Sickles. Digital mammography : Roc studies of the effects of pixel size and unsharpmask filtering

- on the detection of subtle microcalcifications. *Investigative Radiology*, 22(7) :581–589, 1987.
- [48] K. J. Bovis and S. Singh. Enhancement technique evaluation using quantitative measures on digital mammograms. *Proceedings of 5th International Workshop on Digital Mammography*, pages 547–553, 2001.
- [49] J.M.M. Muñoz, H.J. Ochoa Domínguez, O.O.V Villegas, V.G.C Sánchez, and L.O. Maynez. The nonsampled contourlet transform for enhancement of microcalcifications in digital mammograms. *Springer-Verlag Berlin Heidelberg*, pages 292–302, 2009.
- [50] K. Sankar and K. Nirmala. Nonsampled contourlet transformation based image enhancement with spatial and statistical feature extraction for classification of digital mammogram. *International Journal of Soft Computing and Engineering (IJSCE)*, 3 :292–302, September 2013.
- [51] K. Delac, M. Grgic, and T. Kos. Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions. *International Conference on Systems, Signals and Image Processing IWSSIP*, pages 21–23, September 2006.
- [52] R.C. Gonzalez and R.E. Woods. Digital image processing. *Prentice Hall, Upper Saddle River, NJ*, 2002.
- [53] S.E. Umbaugh. Computer imaging : Digital image analysis and processing. *CRC Press, Florida*, 2005.
- [54] W. H. Liao and D. Y. Li. Homomorphic processing techniques for near-infrared images. *Acoustics, speech, and Signal Processing*, 2003.
- [55] P. Bonnet. Cours de traitement d’image.
- [56] W. Eziddin. Segmentation itérative d’image par propagation de connaissances dans le domaine possibiliste : Application à la détection de tumeurs en imagerie mammographique. *Thèse. Université de Bretagne Occidentale*, Juin 2012.
- [57] F. Eddaoudi. Aide la détection des pathologies du sein par les techniques de segmentation et de codage de textures des images mammographiques. *Thèse de doctorat Faculté des Sciences de Rabat*, 2012.
- [58] M. Kallergi, K. Woods, L. Clarke, W. Qian, and R. Clark. Image segmentation in digital mammography comparison of local thresholding and region growing algorithms. *Computerized medical imaging and graphics*, 16(5) :231–331, 1992.
- [59] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Shmidt. Analysis of spiculation in the computerized classification of mammographic masses. *Medical Physics*, 22(10) :1569–1579, 1995.
- [60] S. L. Horowitz and S. Pavlidis. Picture segmentation by a directed split and merge procedure. *In Second International Joint Conference on Pattern Recognition*, pages 424–433, 1974.
- [61] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6) :679–698, 1986.
- [62] J. M. S. Prewitt. Object enhancement and extraction. picture processing and psychopictorics. *Academic Press, New York*, 1970.

- [63] I. E. Sobel. Camera models and machine perception. *Thèse de Doctorat, Stanford University*, 1970.
- [64] L. G. Roberts. Machine perception of three-dimensional solids. *Optical and Electrooptical Information processing*, pages 159–197, 1963.
- [65] M. Kass, A. Witkin, and D. Terzopoulos. Snakes : active contour models. *International Journal of Computer Vision*, 1(4) :321–331, 1987.
- [66] D. Terzopoulos and K. Fleischer. Deformable models. *Visual Computer*, 4(6) :306–331, 1988.
- [67] S. Beucher. The watershed transformation applied to image segmentation. *Centre de Morphologie Mathématique, Ecole des Mines de Paris*.
- [68] J. Cousty. Lignes de partage des eaux discrètes :théorie et application à la segmentation d’images cardiaques. *Thèse. Université de Marne-La-Vallée, Paris*, 2007.
- [69] J. Métivier. Radiothérapie externe du sein : amélioration de l’uniformité de la dose au volume cible. *Mémoire présenté à l’Université Laval pour l’obtention du grade de Maître Sciences (M.Sc.)*, Sep 2003.
- [70] T.C. Wang and N.B. Karayiannis. Detection of microcalcification in digital mammograms using wavelets. *IEEE Trans. Med. Imaging*, 17 :498–509, 1998.
- [71] R.N. Strickland and H.I. Hahn. Wavelet transform for detecting microcalcification in mammograms. *IEEE Trans. Med. Imaging*, 15, 1996.
- [72] M. Alolfe, W. Mohamed, A. Youssef, Y. Kadah, and A. Mohamed. Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms. *NRSC*, Mar 2009.
- [73] N. Hamad, M. Ellauze, and M. Bouhlel. Wavelets investigation for computer aided detection of microcalcification in breast cancer. *IEEE Xplore Digital Library*, 2009.
- [74] T. Balakumaran, Dr.ILA. Vennila, and C. Gowri Shankar. Detection of microcalcification in mammograms using wavelet transform and fuzzy. *International Journal of Computer Science and Information Security IJCSIS*, 7(1), 2010.
- [75] H.D. Cheng, J. Wang, and X. Shi. Microcalcification detection using fuzzy logic and scale space approaches. *Pattern Recognition*, 37(2) :363 – 375, 2004.
- [76] H. Chan, K. Doi, S. Galhotra, C.J. Vyborny, H. MacMahon, and P.M. Jokich. Image feature analysis and computer-aided diagnosis in digital radiography. i. automated detection of microcalcifications in mammography. *Medical physics*, pages 538–548, 1987.
- [77] T. Stojic and B. Reljin. Enhancement of microcalcifications in digitized mammograms : Multifractal and mathematical morphology approach. *FME Transactions*, 38(1) :538–548, 2010.
- [78] S. Al-Kindi and G. Al-Kindi. Breast sonogram and mammogram enhancement using hybrid and repetitive smoothing-sharpening technique. *IEEE Xplore Digital Library*, 2011.
- [79] A.B. Huddin, Brian W.-H. Ng, and Derek Abbott. Investigation of multiorientation and multiresolution features for microcalcifications classification in mammograms. *IEEE Xplore Digital Library*, 2011.

- [80] M.M. Eltoukhy, I. Faye, and B.B. Samir. Breast cancer diagnosis in digital mammo-gram using multiscale curvelet transform. *Comput. Med. Imag. Graphics*, 34 :269–276, 2010.
- [81] M.M. Eltoukhy, I. Faye, and B.B. Samir. A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram. *Computers in Biology and Medicine*, 40 :384,391, 2010.
- [82] J.S. Leena Jasmine, S. Baskaran, and A. Govardhan. A robust approach to classify mi-crocalcification in digital mammograms using contourlet transform and support vector machine. *American Journal of Engineering and Applied Sciences*, pages 384,391, 2013.
- [83] T. Bhangale, U. B. Desai, and U. Sharma. An unsupervised scheme for detection of microcalcifications on mammograms. *IEEE Xplore Digital Library*, 2000.
- [84] R. Lakshmanan and T. Vinu. Microcalcification detection by morphology,singularities of contourlet transform and neural network. *Bonfring International Journal of Net-working Technologies and Applications*, 1(1), 2012.
- [85] A. Pujol. Contribution à la classification sémantique d’image. *thèse. École Centrale de Lyon*, Juin 2009.
- [86] L. Vincent. Texture segmentation using gabor filters, center for intelligent machines, mcgill university. décembre 2000.
- [87] A. Poursaberi. Wavelet-based recognition of facial expressions and faces. *Thesis submitted to the faculty of graduate studies in partial fulfilment of the requirements for the degree of doctor of philosophy*, September 2014.
- [88] A. Poresbki. Sélection d’attributs de texture couleur pour la classification. Novembre 2009.
- [89] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41 :909–996, 1988.
- [90] S. Mallat. A theory of multiresolution signal decomposition : the wavelet representa-tion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 :674–693, July 1989.
- [91] Y. Meyer. Principe d’incertitude, bases hilbertiennes et algèbres d’opérateurs. *Séminaire Bourbaki*, pages 1985–1986.
- [92] A. Grossman and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4) :723–736, 1984.
- [93] I. Daubechies. Ten lectures on wavelets. *Society for Industrial and Applied Mathema-tics, Philadelphia, PA, USA*, 1992.
- [94] S. Mallat. A wavelet tour of signal processing. *Academic Press, San Diego, 2nd edition*, 1999.
- [95] Y. Meyer. Ondelettes et opérateurs. *book*, 1990.
- [96] M. Vetterli. Multi-dimensional sub-band coding : Some theory and algorithms. *Signal Processing*, 6 :97–112, 1984.
- [97] J. Landré and F. Truchetet. Boite à outils ondelettes sous matlab. *Université de Bourgogne formation de C.R.I*, Mars 2002.

- [98] P. S. Addison. The illustrated wavelet transform handbook. *Institute of physics publishing, Bristol and Philadelphia, UK*, 2002.
- [99] S. Grgic, M. Grgic, and B. zoko cihlar. performance analysis of image compression using wavelets. *IEEE Trans, industrial electronics*, 48(3), Juin 2001.
- [100] S. Benchikh. Transformées orthogonales de l’analyse spectrale pour le filtrage et la compression des images. *Mémoire du diplôme de maitrise (M.Sc.A)*, Avril 2011.
- [101] S. Mallat. A theory for multiresolution signal decomposition : the wavelet decomposition. *IEEE Trans*, 11(7) :674–693, July 1989.
- [102] M.N. Do and M. Vetterli. Pyramidal directional filter banks and curvelets. *IEEE International Conference on Image processing, Thessaloniki, Greece*, Oct 2001.
- [103] M.N. Do and M. Vetterli. Contourlets : a directionmultiresolution image representation. *IEEE In on Image processing, Rochester, USA*, Sept 2002.
- [104] A.L. Cunha, J. Zhou, and M.N. Do. The nonsubsampling contourlet transform : Theory, design, and applications. *IEEE Trans. Image Process*, 15(10) :3089–3101, Oct 2006.
- [105] D. Ghanima. Comparatif de méthodes avancées pour la fusion d’images satellites. *Thèse. University of Sciences and Technology of Oran*, 2015.
- [106] M.J. Homer. Mammographic interpretation : A practical approach. *The Radiologic Clinics of North America, Breast Imaging*, 1997.
- [107] N. Karssemeijer. Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology*, pages 365–378, 1998.
- [108] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, and A.F. Frère. Segmentation of mammograms : Identification of the skin - air boundary, pectoral muscle, and fibroglandular disc. *Proceedings of the 5th International Workshop On Digital Mammography, Toronto, Canada, 11-14 June 2000, Medical Physics Publishing*, pages 573–579, June 2001.
- [109] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, and A.F. Frère. Automatic identification of the pectoral muscle in mammograms. *IEEE Transactions on Medical Imaging*, pages 232–245, June 2004.
- [110] H. Mirzaalian, M.R. Ahmadzadeh, and S. Sadri. Pectoral muscle segmentation on digital mammograms by nonlinear diffusion filtering. *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C. Canada, 22-24*, pages 581–584, August 2007.
- [111] M. Mustra, J. Bozek, and M. Grgic. Breast pectoral muscle detection using wavelet decomposition. *IEEE*, 2009.
- [112] A. boucher, P. Jouve, F. Cloppet, and N. Vincent. Segmentation du muscle pectoral sur une mammographie. *ORASIS’09 - Congrès des jeunes chercheurs en vision par ordinateur Trégastel, France*, 2009.
- [113] K.S. Camilus, V.K. Govindan, and P.S. Sathidevi. Pectoral muscle identification in mammograms. *Computer-aided identification of the pectoral muscle in digitized mammograms- Journal of Digital Imaging*, 23(5), October 2010.

- [114] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and M.T. Rickard. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Trans Med Imaging*, 2004.
- [115] D. Raba, A. Oliver, J. Mart, M. Peracaula, and J. Espunya. Breast segmentation with pectoral muscle suppression on digital mammograms. *Lect Notes Comp Sci*, 2005.
- [116] F. Ma, M. Bajger, and M.J. Bottema. Extracting the pectoral muscle in screening mammograms using a graph pyramid. *APRS workshop on digital image computing, University of Queensland*, 2005.
- [117] J. Dengler, S. Behrens, and J.F. Desaga. Segmentation of microcalcification in mammograms. *IEEE Transactions on Medical Imaging*, 12(4) :634–642, 1993.
- [118] I. Bankman, J. Tsai, D. Kim, O. Gatewood, and W. Brody. Detection of microcalcification clusters using neural networks. *IEEE Engineering in Medicine and Biology Society*, pages 590–591, 1994.
- [119] H. D. Cheng, X. Cai, X. Chen, L. Hu, and X. Lou. Computer-aided detection and classification of microcalcifications in mammograms. *a survey. Pattern Recognition*, 36(12) :2967–2991, 2003.
- [120] J. S. Weszka, C. R. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6(4) :269–285, 1976.
- [121] M. M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4(2) :172–179, 1975.
- [122] J. K. Kim and H.W. Park. Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3) :231–238, 1999.
- [123] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6) :610–621, 1973.
- [124] D. Wei, H.P. Chan, N. Petrick, B. Sahiner, M.A. Helvie, D. D. Adler, and M. M. Goodstitt. False-positive reduction technique for detection of masses on digital mammograms : Global and local multiresolution texture analysis. *Medical Physics*, 24(6) :903–914, 1997.
- [125] R. Llobet, R. Paredes, and J. C. Pérez-Cortes. Comparison of feature extraction methods for breast cancer detection. *Iberian Conference on Pattern Recognition and Image Analysis (PRIA)*, pages 495–502, 2005.
- [126] N. Mudigonda, R. Rangayyan, and J. Leo Desautels. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging*, 20(12) :1215–1227, 2001.
- [127] S. Beura, S. Majhi, and R. Dash. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing*, 20(12) :1215–1227, 2015.
- [128] K. Houari. Recherche d’image par le contenu. *Thèse. Université MENTOURI CONSTANTINE*, Juin 2010.
- [129] H. Bannour. Une approche sémantique basée sur l’apprentissage pour la recherche d’image par contenu. *Conférence en Recherche d’Information et Applications*, 2009.

- [130] B. Gregory. Développement d'outils de navigation hypermédia (images) pour dispositifs mobiles(android). 2011.
- [131] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of computer.*, 1991.
- [132] M. Stricker and M. Orengo. Similarity of color images. *SPIE*, 1995.
- [133] M. Hanifi. Extraction de caractéristiques de texture pour la classification des images satellite. 2009.
- [134] M.K. Hu. Visual pattern recognition by moments invariants. *IRE TRANSACTIONS ON INFORMATION THEORY*, 1962.
- [135] Z. Zaabot. Les réseaux bayésiens : Application en reconnaissance de forme à partir d'information complète ou incomplète. *Mémoire de magistère. Université MOULOUD MAMMERI, TIZI-OUZOU*, 2012.
- [136] A. Poresbki. Sélection d'attributs de texture de texture couleur pour la classification. *Thèse. Université LILLE*, Novembre 2009.
- [137] N. Idrissi. La navigation dans des bases d'images : prise en compte des attributs de texture. *thèse doctorat. Université de Nantes*, Octobre 2008.
- [138] S. Bedouhene. Recherche d'image par le contenu. *Mémoire magister. Université MOULOUD MAMMERI, TIZI-OUZOU*, Octobre 2008.
- [139] B. Gregory. Développement d'outils de navigation hypermédia (images) pour dispositifs mobiles(android). *Mémoire de licence. Université de Mons*, 2011.
- [140] H. Tamura, S. Mori, and T. Yamawaki. Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics, SMC*, 8(6) :460–473, 1978.
- [141] A. Znaidia, T. Zaharia, and F. Preteux. Une évaluation des descripteurs visuels mpeg-7 pour la recherche d'images par le contenu. *Institut Télécom ; Télécom SudParis ; Département ARTEMIS*.
- [142] P. Techasith. Image search engine. *Project report, Imperial College, London, UK*, juillet 2002.
- [143] P. Howarth and S. Rüger. Evaluation of texture features for content-based image retrieval. *In Proceedings of the International Conference on Image and Video Retrieval (CIVR'04)*, 3115 :326–334, juillet 2004.
- [144] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features of image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), Nov 1973.
- [145] H. Bennour. Indexation automatique des images : une approche sémantique basé sur l'apprentissage supervisé de la régularité. *Conférence en Recherche d'Information et Applications. CORIA*, 2009.
- [146] C. Lu, P. Chung, and C. Chen. Unsupervised texture segmentation via wavelet transform. *Pattern Recognition, Elsevier*, 30(5) :729–742, 1997.
- [147] R. Soulard and P. Carré. Ondelettes quaternioniques pour la classification de textures. *17ème Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA'10)*, pages 222–229, 2010.



- [148] A. Torralba, K.P. Murphy, and W.T. Freeman. Sharing visual features for multiclass and multiview object detection. *TPAMI*, 2007.
- [149] A. Oliva and A. Torralba. Modeling the shape of the scene : A holistic representation of the spatial envelope. *International Journal of Computer Vision*, pages 145–175, 2001.
- [150] A. Oliva and A. Torralba. Building the gist of a scene : the role of global image features in recognition, progress in brain research. *International Journal of Computer Vision*, pages 1–19, 2006.
- [151] Hans G. Feichtinger and T. Strohmer.abor analysis and algorithms. *Birkhäuser*, pages 1–19, 1998.
- [152] O. Augereau and N. Journet. Reconnaissance et extraction de pièces d’identité : Une application industrielle à la détection de cartes d’identité et de passeports. *1re soumission à CIFED 2012*, pages 1–19, décembre 2011.
- [153] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1) :21–27, 1967.
- [154] W. J. Veldkamp and N. Karssemeijer. Improved method for detection of microcalcification clusters in digital mammograms. *SPIE Symposium on Medical Imaging*, 3661 :512–522, 1999.
- [155] H. Zadeh, S. Nezhad, and F. Rad. Shape-based and texture-based feature extraction for classification of microcalcification in mammograms. *SPIE Symposium on Medical Imaging*, 4684 :301–310, 2001.
- [156] D. Rumelhart, G. Hinton, and R. Williams. Learning representation by back-propagating errors. *Nature Publishing Group*, 323(9) :533–536, 1986.
- [157] C. Lau. Neural networks : Theoretical foundations and analysis. *IEEE Press, Piscataway, NJ, USA*.
- [158] C. Bourrely and S. Muller. Detection of microcalcifications in mammographic images. *NATO ASI Series in Computer and System Sciences*, 68 :325–328, 1989.
- [159] Y. Chitre, A. Dhawan, and M. Moskowitz. Artificial neural network based classification of mammographic microcalcifications using image structure and cluster features. *IEEE Engineering in Medicine and Biology Society*, 68 :592–593, 1994.
- [160] I. Bankman, J. Tsai, D. Kim, O. Gatewood, and W. Brody. Detection of microcalcification clusters using neural networks. *IEEE Engineering in Medicine and Biology Society*, pages 590–591, 1994.
- [161] J. K. Kim and H.W. Park. Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3) :231–238, 1999.
- [162] D. Kramer and F. Aghdasi. Texture analysis techniques for the classification of microcalcifications in digitised mammograms. *IEEE AFRICON*, pages 395–400, 1999.
- [163] J. Zheng and E. Regentova. Cad system for detecting clustered microcalcifications in digital mammograms using independent component analysis and neural network. *International Journal of Computer Assisted Radiology and Surgery*, pages 517–527, 2006.

- [164] V. N. Vapnik. The nature of statistical learning theory. *Book. Springer-Verlag New York*, 1995.
- [165] A. Cao, Q. Song, X. Yang, S. Liu, and C. Guo. Mammographic mass detection by vicinal support vector machine. *IEEE International Joint Conference on Neural Networks*, pages 1953–1958, 2004.
- [166] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli. A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Physics in Medicine and Biology*, 49(6) :961–975, 2004.
- [167] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern classification. *Book*, 2000.
- [168] D. Parrochia. Classifications, histoire et problèmes formels. cinquièmes rencontres de la société francophone de classification. *SFC'97*, pages 13–18, Septembre 1997.
- [169] V.N. Vapnik. The nature of statistical learning theory. *Book. Springer-Verlag, New York, NY*, 1995.
- [170] H. Mohamadally and B. Fomani. Svm : Machines à vecteurs de support ou séparateurs à vastes marges. *Versailles St Quentin, France*, janvier 2006.
- [171] N. Marref. Apprentissage incrémental and machines à vecteurs supports. *Mémoire Magister. Université HADJ LAKHDAR-BATNA Faculté des sciences Département d'Informatique*, 2013.
- [172] J. Callut. Implementation efficace des support vector machines pour la classification. *Master thesis. Université libre de Bruxelles*, 2003.
- [173] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1) :21–27, 1967.
- [174] K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, and N. Kerdprasop. An empirical study of distance metrics for k-nearest neighbor algorithm. *Proceedings of the 3rd International Conference on Industrial Application Engineering*, 2015.
- [175] S. Haykin. Neural networks - a comprehensive foundation. *Macmillan College Publishing Company, New York*, 1994.
- [176] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5 :115–133, 1943.
- [177] F. Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological*, 65(6), 1958.
- [178] B. Widrow and M.A. Lehr. Perceptrons, adalines and backpropagation. *The Handbook of Brain Theory and Neural Networks*, pages 719–724, 1958.
- [179] L. Houam. Contribution à l'analyse de texture de radiographies osseuses pour le diagnostic précoce de l'ostéoporose. *Thèse. Université d'Orléans*, 2013.
- [180] K. Veropoulos. Machine learning approaches to medical decision making. *PhD thesis, University of Bristol, United Kingdom*, 2001.
- [181] S Jae. Lim. Two-dimensional signal and image processing. *NJ : Prentice Hall*, pages 213–217, 1990.

- [182] M. Wirth and D. Nikitenko. Suppression of stripe artifacts in mammograms using weighted median filtering. *Chapter Image Analysis and Recognition of the series Lecture Notes in Computer Science*, 3656 :966–973, 2005.
- [183] S. Boucher. The watershed transformation applied to image segmentation. *Centre de morphologie mathématique*, 1992.