

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY

UNIVERSITE BADJI MOKHTAR-ANNABA



جامعة باجي مختار - عنابة

Faculté des Sciences de l'Ingénierat
Département d'Informatique

Année : 2019/2020

THESE

Présentée en vue de l'obtention du diplôme de

Doctorat 3^{ième} Cycle LMD

Intitulée :

Systeme basé apprentissage et super-résolution pour guider le diagnostic du cancer du sein

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : TIVA (Traitement d'Image et Vision Artificielle)

BOUDRAA SAWSEN

Devant le jury :

Bouhadada Tahar	Prof.	Université Badji Mokhtar – Annaba	Président
Merouani Hayet Farida	Prof.	Université Badji Mokhtar – Annaba	Directrice de thèse
Bendjenna Hakim	Prof.	Université de Tebessa	Examineur
Azizi Nabiha	Prof.	Université Badji Mokhtar – Annaba	Examineur
Melouah Ahlem	MCA.	Université Badji Mokhtar – Annaba	Invitée

Dédicace

A ma chère mère,

A mon cher père,

A mon cher Bilel,

A mes frères et sœurs,

A toute ma famille,

A mes amis,

A mes collègues.

Remerciements

Avant toute chose je tiens à remercier **Allah** pour l'accomplissement de ce projet.

Ma haute gratitude, mes profonds respects, et mes sincères remerciements à ma directrice de thèse madame **Hayet Farida Merouani**, professeur à l'université Badji Mokhtar d'Annaba, pour sa présence remarquable durant la réalisation de ce travail et son engagement dans toutes les étapes de mon encadrement, pour sa patience et ses judicieux conseils, et pour sa disponibilité tout au long de ces cinq années.

Je tiens à remercier très sincèrement docteur **Melouah Ahlem**, pour ses encouragements, son soutien, ses aides, et ses éclairages techniques. Elle a su m'intégrer dans le domaine réel de la recherche, me pousser à toujours chercher le mieux, elle fût tout le temps disponible malgré ses nombreuses charges. Je vous suis reconnaissante d'avoir veillé au bon déroulement de mon projet et de m'avoir permis de bénéficier de vos conseils et orientations scientifiques toujours très constructifs.

J'exprime ma profonde reconnaissance et mes chaleureux remerciements aux membres de jury qui ont accepté de juger ce travail et d'y apporter leur caution :

- Monsieur **Bouhadada Tahar**, Professeur à l'université Badji Mokhtar d'Annaba.
- Monsieur **Bendjenna Hakim**, Professeur à l'université de Tebessa.
- Madame **Azizi Nabiha**, Professeur à l'université Badji Mokhtar d'Annaba.
- Madame **Melouah Ahlem**, Maitre de conférences classe A à l'université Badji Mokhtar d'Annaba.

J'adresse mes vifs remerciements à madame **Souici Labiba**, Professeur à l'université Badji Mokhtar d'Annaba, pour ses orientations scientifiques et pour la pertinence de ses conseils.

Merci à toute l'équipe du laboratoire LRI (Laboratoire de Recherche en Informatique).

Je tiens aussi à remercier le personnel du département de l'informatique de l'université Badji Mokhtar Annaba et tous les enseignants qui, par leurs enseignements, leurs encouragements et leurs aides, ont contribué à ma formation.

Résumé

Le cancer du sein est un véritable problème de santé public. Plusieurs femmes atteintes de cette maladie sont décédées. Le cancer du sein est parmi les cancers les plus mortels.

Actuellement, le seul moyen pour lutter contre ce fléau est la détection précoce de masse mammaire. Le diagnostic assisté par ordinateur est un grand secours dans ce sens.

Un système de diagnostic assisté par ordinateur analyse automatiquement les mammographies numériques pour déceler les zones suspectes et offrir aux experts un deuxième avis. Ce deuxième avis est le résultat d'un processus automatique basé sur une importante étape : l'extraction des caractéristiques. Améliorer les performances des étapes d'extraction des caractéristiques et de classification est l'objectif de cette thèse.

Cinq combinaisons de caractéristiques de texture du premier, deuxième et troisième ordre ont été testées. La combinaison des descripteurs du premier et troisième ordre a été testée sur plusieurs images mammographiques provenant de la base d'images MIAS, et elle a donné le meilleur résultat.

En plus, pour renforcer la discrimination entre les masses malignes et les masses bénignes nous avons ajouté une étape de "super-résolution". La super-résolution est une transformation de l'image de sorte à renforcer les spécificités de cette dernière.

L'augmentation de la résolution des images mammographiques par cette technique permet une meilleure représentation et fournit des détails capables de différencier les deux types de masses.

Les tests élaborés ont permis de confirmer que la combinaison des caractéristiques du premier et troisième ordre extraites d'une mammographie qui a subi une transformation par super-résolution a fourni le meilleur taux de classification qui avoisine 96%.

Mots clés : Système d'aide au diagnostic, Classification, Cancer du sein, Image mammographique, Descripteurs de texture, Super-résolution.

Abstract

Breast cancer is a real public health problem. Several women with this disease have died. Breast cancer is one of the most deadly cancers.

Currently, the only way to fight against this scourge is the early detection of breast mass. The computer-assisted diagnosis is a great help in this context.

A computer assisted diagnosis system automatically analyzes digital mammograms to detect suspicious areas and offers experts a second opinion. This second opinion is the result of an automatic process based on an important step : feature extraction. Improving the performance of the feature extraction and classification steps is the central subject of this thesis.

Five combinations of first, second and third order texture features have been tested. The combination of the first and third order descriptors was tested on several mammographic images from the MIAS database, and it led to the best result.

In addition, we added a "super resolution" step in order to enhance the discrimination between malignant and benign masses. Super-resolution is a transformation of the image which aims to reinforce its specificities.

Increasing the resolution of mammographic images by this technique allows a better representation and provides details capable of differentiating the two types of masses.

The elaborated tests have confirmed that the combination of the first and the third-order features extracted from a mammogram that underwent super-resolution transformation provided the best classification rate of around 96%.

Keywords : Diagnosis aid system, Classification, Breast cancer, Mammographic image, Texture descriptors, Super-resolution.

الملخص

يشكل سرطان الثدي مشكلة صحية عامة حيث قاد إلى وفاة أعداد مهمة من النساء المصابات به. يعد سرطان الثدي من بين أكثر أنواع السرطانات فتكًا.

حاليًا، يشكل الكشف المبكر عن كتل الثدي الطريقة الوحيدة لمحاربة هاته الأفة. و في هذا السياق، فإن التشخيص بمساعدة الكمبيوتر يقدم مساعدة مهمة و حاسمة.

يقوم نظام التشخيص بمساعدة الكمبيوتر بتحليل صور الثدي الشعاعية الرقمية آليا لتحديد المناطق المشبوهة وتزويد الخبراء برأي ثان. يتمكن النظام من اقتراح هذا الرأي عن طريق تنفيذ نسق آلي يستند على مرحلة مهمة ألا وهي استخلاص المميزات. الهدف من هذه الأطروحة هو تحسين أداء المرحلتين 'استخلاص المميزات و التصنيف'.

تم اختبار خمس تراكيب لمميزات النسيج من الدرجات الأولى والثانية والثالثة. تم اختبار تركيب المميزات من الدرجة الأولى والثالثة على العديد من صور التصوير الشعاعي للثدي من قاعدة بيانات صور MIAS ، وقد أعطى أفضل نتيجة.

إضافة لما سبق، فقد قمنا بإضافة خطوة "الدقة الفائقة" من أجل تعزيز التمييز بين الكتل الخبيثة و الحميدة. حيث تنتج هاته التقنية تحويلا سوريا معززا لتفاصيل الصور.

تسمح زيادة دقة الصور الشعاعية للثدي بهذه التقنية بتمثيل أفضل وتوفر تفاصيل قادرة على التمييز بين نوعي الكتل. أكدت الاختبارات المنجزة بأن الجمع بين مميزات الدرجة الأولى والثالثة المستخرجة من تصوير الثدي الشعاعي الذي خضع لتحويل فائق الدقة يوفر أفضل معدل تصنيف يبلغ حوالي 96٪.

الكلمات المفتاحية : نظام مساعدة التشخيص ، التصنيف ، سرطان الثدي ، الصورة الشعاعية للثدي ، مميزات النسيج ، الدقة الفائقة.

Table des matières

Résumé.....	I
Abstract.....	II
ملخص	III
Table des matières.....	IV
Liste des figures.....	IX
Liste des tableaux	XII
Introduction générale.....	1
1. Contexte et cadre de la recherche	1
2. Problématique et objectif	3
3. Structure de la thèse	3
Partie I : Concepts de base et état de l’art.....	5
Chapitre 01 : Le cancer du sein et la mammographie.....	6
1.1 Introduction	6
1.2 Généralités sur le cancer du sein	6
1.2.1 Définition	6
1.2.2 Incidence et mortalité	8
1.3 Les étiologies du cancer du sein	11
1.4 Les pathologies mammaires	13
1.4.1 Les masses et les calcifications	13
1.4.2 La classification des pathologies mammaires	14
1.4.2.1 Le système BI-RADS (Breast Imaging Reporting and Data System).....	14
1.4.2.2 La classification de LeGal	16
1.5 Les moyens d’exploration pour dépister le cancer du sein.....	16
1.6 Traitements du cancer du sein	22

1.6.1	La chirurgie	22
1.6.2	La radiothérapie.....	22
1.6.3	L'hormonothérapie.....	22
1.6.4	La chimiothérapie.....	23
1.7	Bases de données des images mammographiques numériques	23
1.7.1	La base DDSM	23
1.7.2	La base AMDI.....	24
1.7.3	La base InBreast	24
1.7.4	La base MAGIC-5	25
1.7.5	La base WDBC.....	25
1.8	Conclusion	26
Chapitre 02 : Concepts de base sur la classification des images		27
2.1	Introduction	27
2.2	Système de diagnostic assisté par ordinateur.....	27
2.3	Concepts et méthodes de base pour la classification des images	29
2.3.1	Extraction des caractéristiques	29
2.3.1.1	La notion de texture	31
2.3.1.2	Les attributs statistiques.....	34
2.3.1.3	Attribut basé sur la modélisation spatiale des textures (méthode des motifs locaux binaires)	41
2.3.2	La classification.....	43
2.3.2.1	Les méthodes de classifications non supervisées	45
2.3.2.2	Les méthodes de classification supervisées.....	47
2.3.2.3	Évaluation des algorithmes de classification	49
2.4	La super-résolution	50
2.5	Conclusion	54

Chapitre 03 : Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales	55
---	----

3.1 Introduction	55
3.2 Les systèmes d'aide au diagnostic pour la classification des lésions mammaires	55
3.3 Synthèse des travaux de classification automatique des lésions mammaires	58
3.3.1 Classification basée sur les approches de sélection de caractéristiques	59
3.3.2 Classification basée sur les approches d'apprentissage profond	64
3.4 Application de la super-résolution sur les images médicales	68
3.5 Conclusion	77

Partie II : Contributions.....78

Chapitre 4 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales.....	79
---	----

4.1 Introduction	79
4.2 Ressources utilisées	80
4.2.1 Ordinateur.....	80
4.2.2 Base des images mammographiques.....	80
4.2.3 Environnement de travail	81
4.2.3.1 Eclipse	81
4.2.3.2 NetBeans.....	81
4.2.3.3 Java	81
4.2.3.4 Weka.....	82
4.2.3.5 ImageJ.....	82
4.2.4 Critères d'évaluation	82
4.2.4.1 La précision	83
4.2.4.2 La sensibilité.....	83
4.2.4.3 La spécificité.....	83
4.2.4.4 La matrice de confusion.....	83

4.2.4.5	L'aire sous la courbe ROC	84
4.3	Extraction de l'information de texture par les descripteurs statistiques du premier et du deuxième ordre	84
4.3.1	Hybridation des caractéristiques statistiques.....	85
4.3.1.1	Extraction des caractéristiques	86
4.3.1.2	Classification	89
4.3.1.3	Résultats et discussion	89
4.3.2	Hybridation des descripteurs du premier ordre et de Haralick.....	90
4.3.2.1	Extraction des caractéristiques	91
4.3.2.2	Classification	93
4.3.2.3	Résultats et discussion	93
4.3.3	Classification des masses mammaires basée sur les attributs statistiques du premier ordre.....	96
4.3.3.1	Extraction des caractéristiques	97
4.3.3.2	Classification	97
4.3.3.3	Résultats et discussion	98
4.4	Extraction de l'information de texture par les caractéristiques de Haralick et les motifs binaires locaux.....	99
4.4.1	Représentation profonde de la texture via l'opérateur LBP.....	100
4.4.1.1	La transformation via LBP	101
4.4.1.2	Extraction des caractéristiques	101
4.4.1.3	Classification	102
4.4.1.4	Résultats et discussion	102
4.5	Conclusion	104
Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture.....		
5.1	Introduction	106

5.2	La discrimination des masses mammaires par la super-résolution et les descripteurs de texture	106
5.2.1	Base d'images	107
5.2.2	Application de la super-résolution	108
5.2.3	Extraction des caractéristiques	110
5.2.4	Phase d'apprentissage	114
5.2.5	Classification.....	116
5.2.6	Résultats et discussion.....	117
5.3	Conclusion	119
	Conclusion générale et perspectives.....	120
1.	Récapitulatif.....	120
2.	Expérimentations réalisées	121
3.	Perspectives.....	123
	Annexes	124
	Annexe A « Attributs statistiques de texture ».....	124
	Annexe B « Les algorithmes de classification avec Weka ».....	126
	Annexe C « Implémentation de l'approche proposée »	137
	Références bibliographiques	140
	Références web.....	160
	Liste des publications	161

Liste des figures

Figure 1.1 : Deux types de masses mammaires (Dhungel et al., 2017)	7
Figure 1.2 : Anatomie de la poitrine féminine avec des coupes transversales des lobes et des canaux (web 01)	8
Figure 1.3 : Le taux de mortalité dans le monde (Bray et al., 2018).....	10
Figure 1.4 : Diagramme à barres des taux d'incidence et de mortalité par région, normalisés selon l'âge, pour le cancer du sein féminin en 2018 (Bray et al., 2018).....	11
Figure 1.5 : Les formes des masses (Alharbi et al., 2012).	14
Figure 1.6 : Les contours des masses (Alharbi et al., 2012).	14
Figure 1.7 : La classification standard de Le Gal (Zhang, 2014).	16
Figure 1.8 : Image échographique pour le sein (a) tumeur maligne et (b) tumeur bénigne (Cheikhrouhou, 2012)	17
Figure 1.9 : IRM mammaire (web 03)	17
Figure 1.10 : Tomosynthèse mammaire (web 04).....	18
Figure 1.11 : Images standards d'une mammographie numérique : Les images crânio-caudales (CC, gauche) et médio-latérales obliques (MLO, droite) (web 05)	21
Figure 1.12 : Groupes d'âge des patientes (Tangaro et al., 2008).	25
Figure 2.1 : Processus général d'un SAD dédié à l'image mammographique (Mabrouk et al., 2019).....	28
Figure 2.2 : Les descripteurs des images (Jalalian et al., 2017).	30
Figure 2.3 : Exemple des images ont une texture naturelle (Suresha, 2017).	31
Figure 2.4 : Exemple des images ont une texture régulière artificielle (Suresha, 2017).....	32
Figure 2.5 : Exemple des images ont une texture régulière naturelle (Suresha, 2017).	32
Figure 2.6 : Les différentes classes d'attributs de texture de la littérature (Attia, 2013).	33
Figure 2.7 : Les quatre directions considérées (Porebski, 2009).	36
Figure 2.8 : Exemple de construction de la matrice GLCM (Srinivasan and Shobha, 2008)..	36
Figure 2.9 : Exemple de calcul des matrices GLRLM.	39
Figure 2.10 : Représentation du pixel central g_c et de son voisinage g_p (Paulhac, 2009).	41
Figure 2.11 : Construction d'un motif binaire et calcul du code LBP (Paulhac, 2009).	42
Figure 2.12 : Différentes primitives de texture détectées par le LBP (Mäenpää and Pietikäinen, 2004).....	43
Figure 2.13 : Les techniques de classification (Jalalian et al., 2017).....	44
Figure 2.14 : Les approches de classifications non-supervisées	45

Figure 2.15 : Exemple d'application de la super-résolution (Park et al., 2003).....	51
Figure 2.16 : Modélisation du processus d'acquisition d'images (Yahiaoui, 2019).	52
Figure 2.17 : La super-résolution multi-images (Yahiaoui, 2019).....	52
Figure 2.18 : Classification des techniques de la super-résolution	53
Figure 3.1 : Deux types de masses (a) une masse bénigne et (b) une masse maligne (web 08).	56
Figure 3.2 : Schéma général d'un système d'aide au diagnostic dédié aux images mammographiques	57
Figure 3.3 : Les approches de classification automatique des lésions mammaires.....	58
Figure 3.4 : Taxonomie pour les dimensions des descripteurs de caractéristiques (Krig, 2014).	59
Figure 3.5 : Représentation graphique du processus de la super-résolution (Wallach et al., 2012).....	69
Figure 3.6 : (a) image acquise sans mouvement, (b) image acquise sans mouvement et fusionnée à l'image TDM, (c) image acquise avec des mouvements de rotation et de translation et à haute résolution, (d) image acquise avec des mouvements de rotation et de translation, à haute résolution et fusionnée à l'image TDM (Wallach, 2011).	70
Figure 3.7 : Exemple d'application de la super-résolution sur une image mammographique (a : image à basse résolution ; b : image avec la super-résolution) (Zheng et al., 2010)	71
Figure 3.8 : Exemple d'application de la super-résolution sur l'image PET (a : image originale à basse résolution ; b : image à haute résolution) (Wallach and al., 2012).....	72
Figure 3.9 : Image IRM du genou (a : image originale à BR ; b : image à HR) (Trinh et al., 2014).....	73
Figure 3.10 : Image IRM de la cheville (a : image originale à BR ; b : image à HR) (Trinh et al., 2014).....	73
Figure 3.11 : Résultats du sur-échantillonnage d'une image d'un adulte (a : image originale, b : image à haute résolution) (Shi et al. 2015).....	74
Figure 3.12 : Exemple d'images reconstruites où : (a) et (d) sont des orientations à court axe, (b) et (e) sont des orientations à long axe vertical, et (c) et (f) sont des orientations à long axe horizontal (Odille et al., 2015).	75
Figure 3.13 : Exemple d'image reconstituée à l'aide de la super-résolution (a : image oblique médio-latérale originale, b : image originale de la région d'intérêt, c : image avec SRCNN (Umehara et al., 2017).....	76

Figure 4.1 : Echantillons de la base de données MIAS (a : image bénigne n° 5, b : image bénigne n° 12, c : image maligne n° 28, d : image maligne n° 72).....	81
Figure 4.2 : Exploration des options de Weka	82
Figure 4.3 : Schéma général de l’approche proposée.....	85
Figure 4.4 : Exemple d'une matrice GLCM (Jinan, 2015).....	87
Figure 4.5 : Le processus de la méthode proposée.....	91
Figure 4.6 : Histogramme du taux de classification de la méthode proposée et la méthode de Jehlol et al. (Jehlol et al., 2015).....	95
Figure 4.7 : Schéma de la méthode proposée	96
Figure 4.8 : Les étapes de la méthodologie proposée	100
Figure 4.9 : Exemple d’application de l’opérateur LBP sur une RI.....	101
Figure 4.10 : Histogramme du taux de classification.....	104
Figure 5.1 : Schéma général de la méthode proposée	107
Figure 5.2 : Utilisation du plugin « Parallel Super-Resolution »	108
Figure 5.3 : Résultat de l’application de la SR.....	109
Figure 5.4 : Exemple d’application de la super-résolution : (a) masse bénigne avant SR (b) masse maligne avant SR (c) masse bénigne après SR (d) masse maligne après SR.....	110
Figure 5.5 : Construction de la matrice de longueur de niveau de gris (a : matrice originale, b : matrice GLRLM).....	111
Figure 5.6 : Apprentissage sans l’utilisation de la SR.....	115
Figure 5.7 : Apprentissage avec l’utilisation de la SR	116

Liste des Tableaux

Tableau 1.1 : Les pays qui ont les taux les plus élevés de cancer du sein en 2018 (web 02). ...	9
Tableau 1. 2 : Facteurs de risque liés au cancer du sein dans le monde (Momenimovahed and Salehiniya, 2019).....	13
Tableau 1.3 : Les données probantes pour les recommandations de prise en charge clinique des mammographies par catégorie du système BI-RADS (Eberl et al., 2006).	15
Tableau 1.4 : Modalités d'imagerie pour le dépistage du cancer du sein chez les femmes à risque moyen (Nelson et al., 2016).....	20
Tableau 2.1 : Description des techniques de classification supervisées (Jalalian et al., 2017).	49
Tableau 2.2 : Comparaison entre les techniques de la super résolution du domaine fréquentiel et du domaine spatial (Makwana et Mehta, 2013).	54
Tableau 3.1 : Résumé des travaux de classification des lésions mammaires.	68
Tableau 4.1 : La matrice de confusion	84
Tableau 4.2 : Caractéristiques du premier et du deuxième ordre.....	88
Tableau 4.3 : Exemple de deux vecteurs de caractéristiques	89
Tableau 4.4 : Les métriques d'évaluation utilisées	89
Tableau 4.5 : Résultats de classification obtenus par la méthode proposée et par la méthode de Haralick (Haralick et al., 1973).....	90
Tableau 4.6 : Tableau des formules de caractéristiques rajoutées	92
Tableau 4.7 : Deux exemples de vecteurs de caractéristiques	93
Tableau 4.8 : Métriques d'évaluation utilisées.....	94
Tableau 4.9 : Le taux de précision de la méthode proposée et la méthode de Jehlol et al. (Jehlol et al., 2015).....	94
Tableau 4.10 : Les caractéristiques utilisées	96
Tableau 4.11 : Exemple de deux vecteurs de caractéristiques des images n° 92 et n° 15	97
Tableau 4.12 : Les critères d'évaluation de la troisième méthode	98
Tableau 4.13 : Résultats de classification	99
Tableau 4.14 : Les résultats des matrices de confusion	99
Tableau 4.15 : Exemple de deux vecteurs de caractéristiques des images n°28 et n°107	102
Tableau 4.16 : Métriques d'évaluation.....	103
Tableau 4.17 : Résultats de classification de la méthode proposée et la méthode classique.	103
Tableau 5.1 : Les caractéristiques d'ordre supérieur sélectionnées.	113

Tableau 5.2 : Exemple de deux vecteurs de caractéristiques avec et sans l'utilisation de la super-résolution.	113
Tableau 5.3 : Récapitulatif des performances des différents classifieurs.....	117
Tableau 5.4 : La précision de l'approche proposée par rapport à celle des autres travaux	118

Introduction générale

1. Contexte et cadre de la recherche

Au cours de ces dernières années, la maladie du cancer du sein est considérée comme un véritable problème de santé public dans le monde entier. Il est parmi les cancers les plus fréquents chez les femmes. Rien qu'en 2018, (selon Globocan en 2018) le cancer du sein a touché 2,1 millions de femmes et il a tué 626 000 d'entre elles (Ferlay et al. 2019).

Dans la région du Moyen-Orient et de l'Afrique du Nord (MENA), le cancer du sein représente 31,1 % de l'incidence totale du cancer chez les femmes, avec un taux de mortalité de 20,9 % (Tfaily et al., 2020).

En Algérie, l'incidence du cancer du sein est considérablement augmentée. Le nombre de cas diagnostiqués en 2018 a atteint 11847, soit 24% des cas d'incidence du cancer chez les femmes algériennes (Tfaily et al., 2020). Malheureusement, en Algérie il n'existe pas de programme solide pour le dépistage du cancer du sein induisant à un diagnostic tardif de cette maladie. Le traitement devient alors lourd et coûteux et parfois inefficace ce qui conduit à un taux de mortalité croissant.

La détection précoce du cancer du sein est primordial, afin d'éviter les traitements compliqués et de réduire les taux de morbidité et de mortalité. Plusieurs organisations appellent au dépistage pour les femmes à partir d'un âge qui commence généralement à partir des quarantaines. Le type d'imagerie le plus efficace actuellement pour cette tâche est la mammographie.

La lecture des images mammographiques est une tâche très exigeante pour les radiologues, il s'agit d'un examen répétitif qui exige beaucoup d'attention aux détails, car la grande variabilité de l'apparence des masses peut provoquer des biopsies inutiles ou manquer des masses malignes (Lévy et Jain 2016). La lecture de ce type d'images est également une tâche difficile car le même examen est interprété différemment de temps en temps, selon l'expert. C'est pourquoi, ces dernières années, le développement de méthodes de traitement des images mammographiques est devenu intéressant, afin d'augmenter la précision du diagnostic et de

fournir aux radiologues des conseils supplémentaires et les aider à réduire certains erreurs (Oliveira et al, 2015).

Ces méthodes ont été utilisées pour développer des systèmes de détection assistée par ordinateur et des systèmes de diagnostic assisté par ordinateur (Oliveira et al, 2015).

Les systèmes de détection/diagnostic assistés par ordinateur sont très utiles pour détecter et diagnostiquer les anomalies des seins, ils peuvent servir de seconde vue dans la détection précoce du cancer du sein, car ils aident les médecins à lire et à interpréter l'image mammographique.

La principale différence entre les systèmes de détection et de diagnostic assistés par ordinateur est que les systèmes de détection ne présentent pas les caractéristiques radiologiques des tumeurs mais aident à localiser et à identifier les éventuelles anomalies de l'image et laissent l'interprétation au radiologue, d'autre part, les systèmes de diagnostic aident à caractériser les résultats des images radiologiques identifiés par un radiologue ou un système de détection (Ramadan, 2020).

Un système de détection/diagnostic aide les médecins à améliorer les interprétations des images en termes de précision, de détection, de productivité en temps de lecture, et d'interprétation des images ; l'utilisation de ce système a montré une augmentation des performances des radiologues (Ramadan, 2020).

Les systèmes de détection/diagnostic assistés par ordinateur utilisent des techniques de traitement d'image et de reconnaissance des formes afin de détecter et de classer les anomalies dans les mammographies (Li et al., 2016).

Le cheminement du système d'aide au diagnostic passe par les étapes suivantes : l'étape du prétraitement, l'étape d'extraction de la région d'intérêt (ROI) et l'étape de reconnaissance des formes, où les algorithmes de prétraitement améliorent la qualité des images, en utilisant certaines techniques de prétraitement comme le filtrage, l'amélioration du contraste,...etc ; les méthodes d'extraction de la région d'intérêt impliquent des techniques de segmentation afin de délimiter la zone de masse.

Dans ce contexte, notre étude portera essentiellement sur le processus de reconnaissance des formes qui comprend deux actions : l'extraction des caractéristiques et la classification des zones d'intérêt dans les images mammographiques. Dans la première action, toutes les informations discriminantes seront calculées à partir des masses mammaires cernées et les caractéristiques significatives sont sélectionnées, et le résultat de cette étape sera un vecteur

de valeurs de caractéristiques. Dans l'action de classification, un classifieur génèrera une décision sur la malignité de la lésion suspecte en utilisant les descripteurs sélectionnés.

2. Problématique et objectif

La reconnaissance des formes est l'étape la plus difficile dans un système d'aide au diagnostic, car il est difficile de distinguer une masse bénigne d'une masse maligne, même pour un expert.

L'extraction des caractéristiques est l'étape clé dans les systèmes d'aide au diagnostic, elle décrit la masse mammaire ; le mauvais choix des caractéristiques dégrade les performances du système et il implique une classification erronée, c'est pourquoi le bon choix des caractéristiques est très important, il garantit un bon taux de classification des masses.

Un grand nombre de méthodes sont proposées dans la littérature afin de trouver la bonne combinaison de caractéristiques qui est fondée généralement sur plusieurs types, tel que le gradient, le niveau de gris, la texture et la forme. Les caractéristiques de texture sont largement utilisées pour l'identification et la discrimination des objets, et plusieurs travaux ont prouvé leurs importances, car elles sont significatives, efficaces et discriminantes. Vu leurs poids, nous allons les intégrer dans tous nos travaux proposés.

Notre problématique se définit comme suit :

- Comment choisir les bons descripteurs de texture ?
- Est ce que l'utilisation de plusieurs types de caractéristique de texture est avantageux ?
- Quel est l'impact de l'ajout d'une nouvelle étape de discrimination ?

L'enjeu de cette thèse est de concevoir un système d'aide au diagnostic robuste capable de discriminer entre les masses bénignes et malignes dans les images mammographiques et de fournir une vue objective aux radiologues dans l'étape de diagnostic.

3. Structure de la thèse

Cette thèse est composée de deux parties :

- **La première partie** expose les notions théoriques fondamentales utilisées et l'état de l'art, elle comprend trois chapitres :
 - Le premier chapitre présente brièvement les notions médicales sur la maladie du cancer du sein (les définitions, l'incidence, la mortalité, les causes, les modalités dédiées au

dépistage, les traitements), il expose également les différentes bases de données des images mammographiques.

- Le second chapitre définit les notions de base sur les étapes clés d'un système d'aide au diagnostique à savoir l'extraction des caractéristiques et la classification des images. Il décrit également les notions liées à la technique de la super-résolution, qui améliore la qualité des images.
- Le troisième chapitre est consacré à la présentation de l'état de l'art des travaux réalisés dans la littérature sur la classification des zones d'intérêt dans les images mammographiques, et sur l'utilisation de la technique de la super-résolution dans le domaine médicale.
- **La deuxième partie** relate les travaux proposés, elle comporte deux chapitres :
 - Le quatrième chapitre expose la première contribution qui comporte quatre expérimentations qui ont pour but la description des masses mammaires via des descripteurs de texture statistiques du premier ordre, du deuxième ordre et les motifs binaires locaux.
 - Le cinquième chapitre aborde notre deuxième contribution, qui a pour objectif l'amélioration de la discrimination des masses mammaires par l'intégration de la méthode de la super-résolution et par la fusion de deux types de caractéristiques de texture.

Enfin, nous clôturons la thèse par une conclusion générale qui présente un résumé de tout ce qui a été réalisé, et nous traçons également les futures perspectives possibles.

Partie I : Concepts de base et état de l'art

Chapitre 01 : Le cancer du sein et la mammographie.

Chapitre 02 : Concepts de base sur la classification des images.

Chapitre 03 : Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales.

Chapitre 01

Le cancer du sein et la mammographie

1.1 Introduction

D'après l'organisation mondiale de la santé, le cancer du sein est l'une des tumeurs malignes les plus fréquents chez les femmes, le nombre de nouveaux cas de ce cancer est toujours en augmentation. La seule opportunité offerte pour lutter contre ce fléau est la détection précoce de la maladie.

Dans le présent chapitre, nous consacrons une grande partie pour présenter les notions médicales sur le cancer du sein telles que, le taux de mortalité, l'incidence dans le monde, les facteurs de risque, les moyens d'exploration dédiés au dépistage de cette maladie, les types de traitements existants et même les bases de données des images mammographiques les plus connues.

1.2 Généralités sur le cancer du sein

1.2.1 Définition

Le cancer du sein est une tumeur maligne provenant des cellules épithéliales des canaux lactifères glandulaires ou des lobules du sein (Kharman-Biz, 2016). Le carcinome du sein est classé comme non invasif ou invasif, selon que la tumeur a commencé ou non à se développer en dehors de la membrane basale. Les carcinomes invasifs sont des cancers dans lesquels les cellules altérées se diffusent dans les tissus conjonctifs environnants et se métastasent dans les organes distants du corps. Environ deux tiers des cancers du sein des carcinomes proviennent des cellules épithéliales des canaux, appelés carcinomes canaux, et environ un tiers des lobules, appelés carcinomes lobulaires (Malhotra et al., 2010 ; Kharman-Biz, 2016).

Il existe deux types de tumeurs : bénignes et malignes (voir figure 1.1). Les tumeurs bénignes ne se propagent pas et n'envahissent pas d'autres parties du corps. De plus, leurs cellules sont souvent plus différenciées et se développent plus lentement que celles des tumeurs malignes. Les fibromes utérins et les grains de beauté sont des exemples de tumeurs bénignes (Zari, 2016).

Les tumeurs bénignes peuvent causer certains dommages tels que des lésions nerveuses, la mort des tissus et la diminution de la circulation sanguine en raison de la compression

provoquée par l'impact des tumeurs sur les tissus ou les organes ; cependant, les tumeurs bénignes ne provoquent généralement pas de maladie mortelle (Zari, 2016).

D'autre part, les tumeurs malignes ont la capacité d'envahir et de migrer du site primaire vers des organes qui résident plus loin du site d'origine via le système lymphatique ou les vaisseaux sanguins, dans un processus appelé métastase. De plus, les cellules malignes ont un taux de prolifération plus élevé, avec un mode de croissance désorganisé, elles sont résistantes à l'apoptose et génétiquement instables. En outre, les tumeurs malignes contribuent souvent à des maladies mortelles (Cooper and Hausman, 2000 ; Zari, 2016).

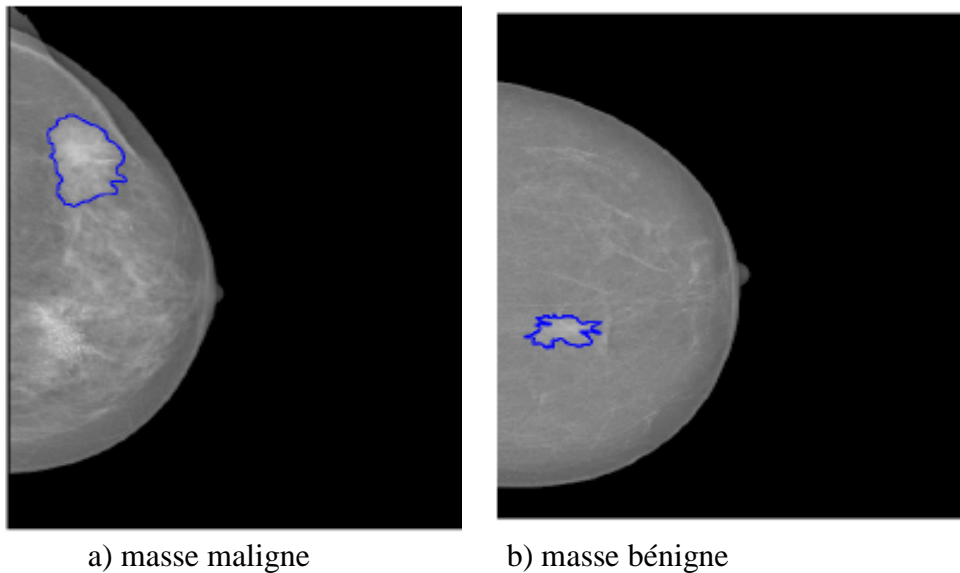


Figure 1.1 : Deux types de masses mammaires (Dhungel et al., 2017)

➤ Anatomie de la poitrine féminine

Le sein féminin est constitué de tissu glandulaire, adipeux et conjonctif répartis en quantités et proportions variables (figure 1.2), ce tissu, lorsqu'il est complètement développé, il comprend 15 à 20 lobes composés de lobules, qui contiennent des groupes d'alvéoles (Ramsay et al., 2005 ; Biermann, 2019).

Les lobes (*lobes*), les lobules (*lobules*) et les alvéoles sont liés par un réseau de conduits convergeant sur le mamelon (*nipple*). Les canaux (*ducts*) et les lobules sont composés de couches de cellules épithéliales et myoépithéliales lumineuses. Au cours de la lactation, les cellules épithéliales lumineuses internes des canaux terminaux et les lobules produisent du lait. Les cellules myoépithéliales externes aident à l'éjection du lait et jouent un rôle dans le maintien de la structure et de la fonction normale du lobule et de la membrane basale. Les autres composants du sein sont les vaisseaux lymphatiques, qui transportent le liquide lymphatique entre les ganglions lymphatiques en formant un réseau dans tout le corps pour

filtrer la lymphe et stocker les globules blancs. Des groupes de ganglions lymphatiques sont situés près du sein, dans l'aisselle, au-dessus de la clavicule et dans la poitrine. De plus, les vaisseaux sanguins et les nerfs se trouvent dans la poitrine (Biermann, 2019).

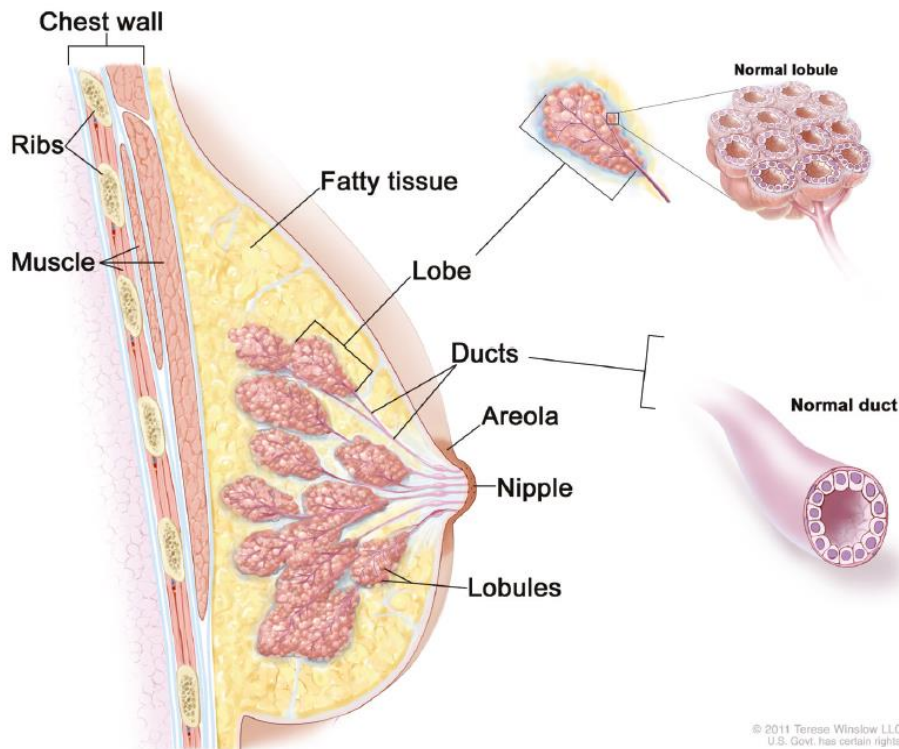


Figure 1.2 : Anatomie de la poitrine féminine avec des coupes transversales des lobes et des canaux (web 01)

1.2.2 Incidence et mortalité

➤ Incidence

Le cancer du sein est considéré comme le cancer le plus fréquent chez les femmes, et c'est le deuxième cancer le plus fréquent dans l'ensemble ; d'après l'institut américain de la recherche sur le cancer, il y a eu plus de 2 millions de nouveaux cas en 2018 (web 02).

Le tableau 1.1 ci-dessous présente les 25 pays ayant les taux les plus élevés du cancer du sein en 2018. Le taux d'incidence du cancer du sein (qui désigne la mesure de la fréquence de nouveaux cas de cancer apparaissant chaque année (Cheikhrouhou, 2012)) augmente régulièrement dans le monde entier et il varie selon quatre régions du monde, où le taux varie de 27 pour 100 000 en Afrique du Moyen-Orient et en Asie de l'Est, et de 92 pour 100 000 en Amérique du Nord (Ferlay et al., 2015 ; Kharman-Biz, 2016). Cela peut être dû à des différences de répartition par âge, de régime alimentaire, de mode de vie, d'ethnicité, de

bagage génétique et d'autres facteurs de risque de cancer du sein entre les populations (Kharman-Biz, 2016).

Rang	Pays	Taux standardisé par âge pour 100,000 femmes
1	Belgique	113.2
2	Luxembourg	109.3
3	Pays-Bas	105.9
4	France (métropolitaine)	99.1
5	Nouvelle-Calédonie (France)	98.0
6	Liban	97.6
7	Australie	94.5
8	Royaume-Uni	93.6
9	Italie	92.8
10	Nouvelle-Zélande	92.6
11	Irlande	90.3
12	Suède	89.8
13	Finlande	89.5
14	Danemark	88.8
15	Suisse	88.1
16	Monténégro	87.8
17	Malte	87.6
18	Norvège	87.5
19	Hongrie	85.5
20	Allemagne	85.4
21	Islande	85.2
22	ÉTATS-UNIS	84.9
23	Canada	83.8
24	Chypre	81.7
25	Samoa	80.1

Tableau 1.1 : Les pays qui ont les taux les plus élevés de cancer du sein en 2018 (web 02).

➤ Mortalité

Selon l'Organisation mondiale de la santé, 8,8 millions de décès dans le monde chaque année sont causés par le cancer ; Le cancer du sein est la première cause de décès dans plus de 100 pays (voir figure 1.3), il représente 13 % de l'ensemble des décès dans le monde (Farooqui

and Ritkia, 2018). La société américaine du cancer (*American Cancer Society*) a déclaré que le taux de mortalité (qui indique le nombre de décès par an (Cheikhrouhou, 2012)) par le cancer du sein chez les femmes a diminué de 39 % entre 1989 et 2015, ces progrès sont attribués à l'amélioration de la détection précoce et de la prévention (Farooqui and Ritkia, 2018).

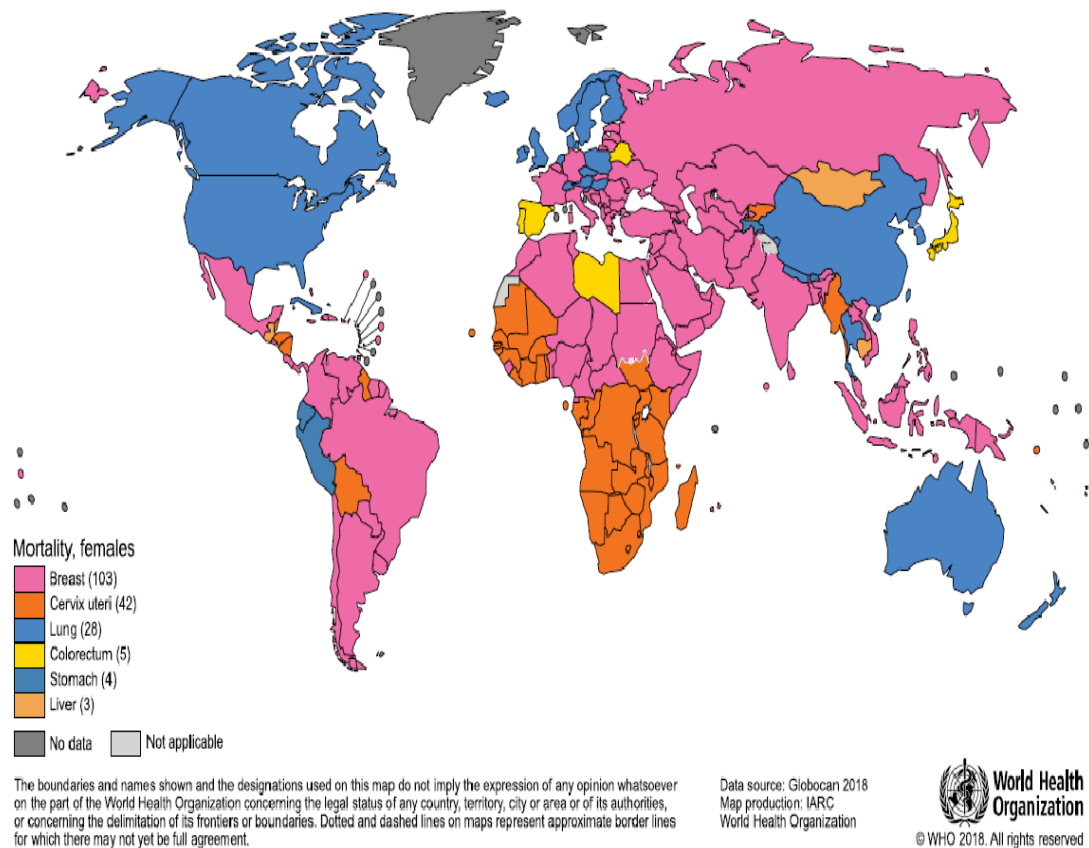


Figure 1.3 : Le taux de mortalité dans le monde (Bray et al., 2018)

Les taux d'incidence du cancer du sein sont les plus élevés en Australie/Nouvelle Zélande, en Europe du Nord (par exemple, au Royaume-Uni, en Suède, Finlande et Danemark), Europe de l'Ouest (Belgique 'avec les taux mondiaux les plus élevés', les Pays-Bas et la France), Europe du Sud (Italie), et l'Amérique du Nord (voir figure 1.4) (Bray et al., 2018).

En termes de mortalité, les taux de cancer du sein présentent une moindre variabilité, avec la plus forte mortalité estimée en Mélanésie, où les Fidji ont les taux de mortalité les plus élevés au monde (Bray et al., 2018).

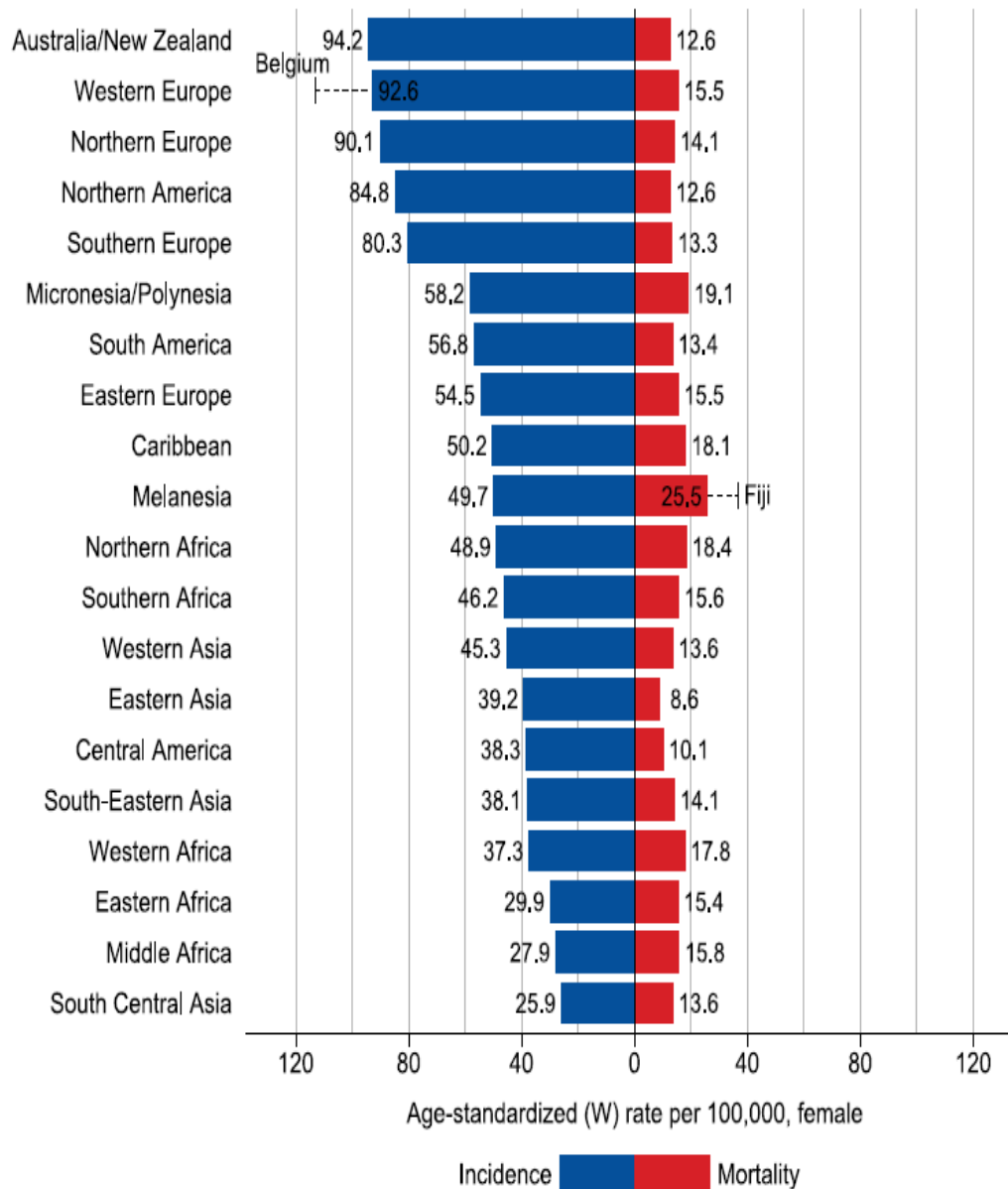


Figure 1.4 : Diagramme à barres des taux d'incidence et de mortalité par région, normalisés selon l'âge, pour le cancer du sein féminin en 2018 (Bray et al., 2018).

1.3 Les étiologies du cancer du sein

Paul et al. (Paul et al., 2015) ont résumé quelques facteurs de risque du cancer du sein tels que : le sexe, l'âge, la race, antécédents familiaux, facteurs génétique, antécédents médicaux personnels comme les ménarches précoces et la ménopause tardive, certaines modifications du génome, la densité des tissus mammaires, l'absence de l'activité physique, la mauvaise alimentation, l'obésité, le manque de sensibilisation, l'alcoolisme, l'exposition aux radiations, les maladies nullipares, la contraception orale et le mode de vie, etc. Dans une étude antérieure, il a été constaté que l'alphabétisation et l'emploi (c'est-à-dire l'enseignement supérieur et le revenu) augmente le facteur de risque de cancer du sein de environ 25 %. Une

Chapitre 01 : Le cancer du sein et la mammographie

étude a révélé que le risque (âge limite) d'être susceptible de contracter un cancer du sein a entre 40 et 50 ans dans les pays asiatiques une décennie plus tôt, mais pour les pays occidentaux, le pic d'âge se situe entre 60 et 70 ans. Autre étude a montré que les femmes non mariées, les femmes nullipares avait un risque de cancer du sein deux fois plus élevé que les multipares femmes et aussi les mariages tardifs (plus de 30 ans) et l'âge de la première grossesse est supérieur à 30 ans (Paul et al., 2015). La plupart des facteurs de risques du cancer du sein sont résumés dans le tableau 1.2.

Facteurs de risque		De Protection	Prédisposer	Controversés
Données démographiques	Genre féminin		✓	
	Âge		✓	
	Groupe sanguin			✓
Reproduction	Âge des ménarches			✓
	L'âge tardif de la ménopause		✓	
	Grossesse à terme	✓		
	Avortement			✓
	Cycle menstruel ovulatoire	✓		
	Caractéristiques de la grossesse	✓	✓	
Hormonal	Méthodes contraceptives hormonales		✓	
	Médicaments stimulant l'ovulation			✓
	Thérapie hormonale post-ménopausique		✓	
Héréditaire	Facteurs génétiques		✓	
	Antécédents familiaux positifs du cancer du sein		✓	
En rapport avec les seins	Réduction de la durée de la lactation	✓		
	Plus de densité mammaire			✓

	Troubles bénins du sein		✓	
Mode de vie	Obésité et surcharge pondérale		✓	
	Consommation d'alcool		✓	
	Fumer		✓	
	Café			✓
	Régime alimentaire		✓	
	Plus d'activité physique	✓		
	Carence en vitamine D	✓		
	Durée du sommeil			✓
Autres	Pollution de l'air		✓	
	Travail de nuit		✓	
	Statut socio-économique		✓	
	Diabète		✓	
	Radiation		✓	

Tableau 1. 2 : Facteurs de risque liés au cancer du sein dans le monde (Momenimovahed and Salehiniya, 2019).

1.4 Les pathologies mammaires

Les radiologues recherchent souvent des signes spécifiques d'anomalies, tels que des grappes de microcalcifications, des masses et des distorsions architecturales sur les mammographies (Alharbi et al., 2012).

1.4.1 Les masses et les calcifications

Selon Sampat et al. (Sampat et al., 2005), une masse est définie comme une lésion occupant un espace, vue dans au moins deux projections différentes (Alharbi et al., 2012).

Les masses peuvent avoir différentes marges (circonscrites, microlobulaires, obscurcies, indéfinie, spiculées) et différentes formes (rondes, ovales, lobulaires, irrégulières), cependant, une distorsion architecturale peut être difficile à déterminer car une masse définie peut être invisible, cela peut se produire sous la forme d'épingles rayonnant à partir d'un point, de rétraction focale ou de distorsion du bord du parenchyme (Alharbi et al., 2012). Les figures 1.5 et 1.6 montrent les différentes formes et contours des masses. Les calcifications se présentent notamment sous la forme de minuscules dépôts de calcium, qui apparaissent comme des taches blanches sur la mammographie (Alharbi et al., 2012).

Le fort contraste et la petite taille ($< 0,5\text{mm}$) caractérisent les microcalcifications par rapport à d'autres éléments, et si leurs taille est supérieure à 1mm , on les appelle des macrocalcifications qui sont souvent bénignes (Cheikhrouhou, 2012).

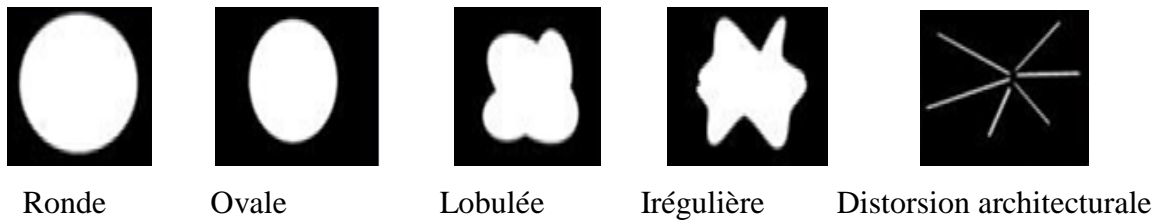


Figure 1.5 : Les formes des masses (Alharbi et al., 2012).

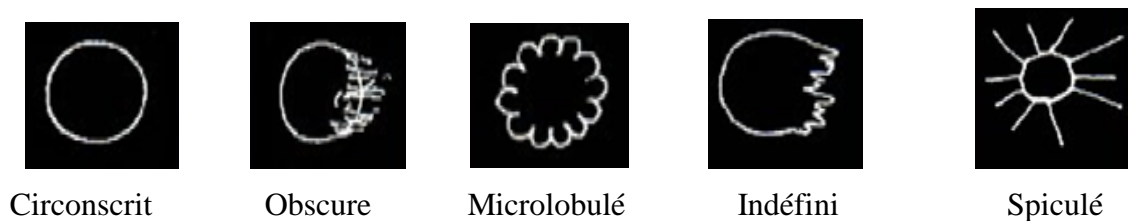


Figure 1.6 : Les contours des masses (Alharbi et al., 2012).

1.4.2 La classification des pathologies mammaires

L'utilisation d'un lexique standard et d'une classification commune offre aux radiologues une description nette des lésions mammaires. Afin d'étudier la morphologie des lésions, nous présentons les deux classifications les plus connues (la classification de BIRADS et la classification de LeGal) (Cheikhrouhou, 2012).

1.4.2.1 Le système BI-RADS (*Breast Imaging Reporting and Data System*)

Le lexique BI-RADS clarifie la communication des résultats des mammographies et il soutient l'achèvement des activités d'amélioration de la qualité et de la recherche clinique (Eberl et al., 2006).

Ce système permet de classifier les images mammographiques en plusieurs catégories en fonction du degré de suspicion de leur caractère pathologique (Cheikhrouhou, 2012).

Les classifications sont divisées en une évaluation incomplète (catégorie 0) et des évaluations achevées (catégories 1, 2, 3, 4, 5, 6), bien qu'il existe 7 catégories d'évaluation, seuls 4 résultats sont possibles : (1) études d'imagerie supplémentaires, (2) mammographie de routine à intervalles réguliers, (3) suivi à court terme, et (4) biopsie. Il a été démontré que les catégories d'évaluation sont corrélées avec la probabilité de malignité, étant donné que chaque

catégorie BI-RADS ne comporte qu'une seule recommandation spécifique, ce système peut à la fois informer les médecins de famille des résultats et orienter le suivi et la prise en charge appropriés (Eberl et al., 2006). Le tableau 1.3 présente les classifications BI-RADS et les recommandations de gestion comme un tableau de preuves.

Catégorie BI-RADS	Évaluation	Recommandation(s) sur la gestion clinique	Force de la recommandation
0	Évaluation incomplète	Nécessité de revoir les études antérieures et/ou de compléter l'imagerie supplémentaire	A
1	Négatif	Continuer le dépistage de routine	A
2	Constataion bénigne	Continuer le dépistage de routine	A
3	Probablement une constataion bénigne	Suivi de la mammographie à court terme (6 mois), puis tous les 6 à 12 mois pendant 1 à 2 ans	B
4	Anomalie suspecte	Effectuez une biopsie, de préférence à l'aiguille	A
5	Fort suspect de malignité ; des mesures appropriées doivent être prises	Biopsie et traitement, si nécessaire	A
6	Malignité connue prouvée par biopsie, traitement en cours	Assurer que le traitement est terminé	

Tableau 1.3 : Les données probantes pour les recommandations de prise en charge clinique des mammographies par catégorie du système BI-RADS (Eberl et al., 2006).

Les caractéristiques mammographiques spécifiques ayant la plus grande valeur prédictive positive de la malignité comprennent les masses avec des marges spiculées et/ou une forme irrégulière, ainsi que les calcifications avec une morphologie linéaire et/ou une distribution segmentée (Eberl et al., 2006).

1.4.2.2 La classification de LeGal

La classification de LeGal (LeGal et al., 1984) englobe cinq groupes identifiant différents types de microcalcifications dans l'ordre croissant du degré de malignité (voir figure 1.7), où le premier groupe décrit les microcalcifications en forme de 'O' et celles partiellement calcifiées, connues sous le nom de microcalcifications en tasse à thé, le deuxième groupe comprend les microcalcifications régulières et rondes de densité uniforme, le troisième est composé de microcalcifications de même forme et de taille inférieure à la seconde, dans le quatrième groupe les microcalcifications irrégulières sont liées au degré élevé de malignité, et le dernier groupe est étroitement lié à un degré très élevé de malignité, ce qu'on appelle la forme vermiculaire (Zhang, 2014).

La classification de LeGal est simple, mais son inconvénient qu'elle se base uniquement sur les microcalcifications (Cheikhrouhou, 2012).

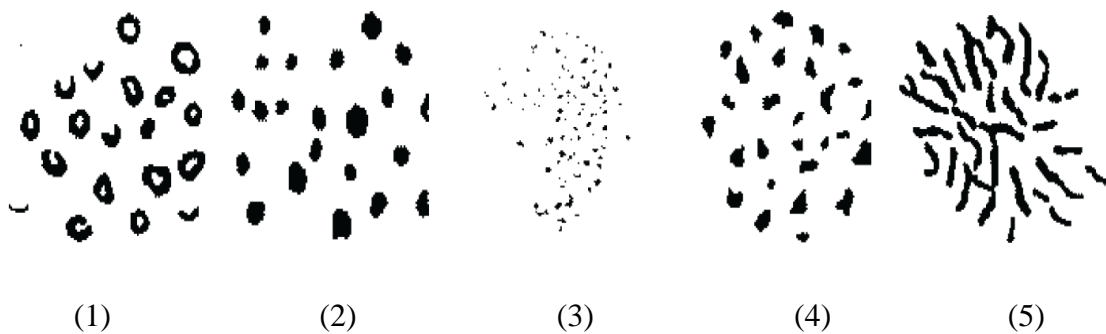


Figure 1.7 : La classification standard de Le Gal (Zhang, 2014).

1.5 Les moyens d'exploration pour dépister le cancer du sein

Pour améliorer les chances de survie et réduire l'utilisation des traitements et des thérapies et, par conséquent, les effets secondaires, de nombreuses modalités d'imagerie sont continuellement mises au point pour diagnostiquer la maladie du cancer du sein le plus tôt possible. Certaines de ces modalités sont utilisées à des fins de dépistage, d'autres à des fins de diagnostic, et quelques autres encore pour une évaluation complémentaire (Subbhuraam et al., 2014).

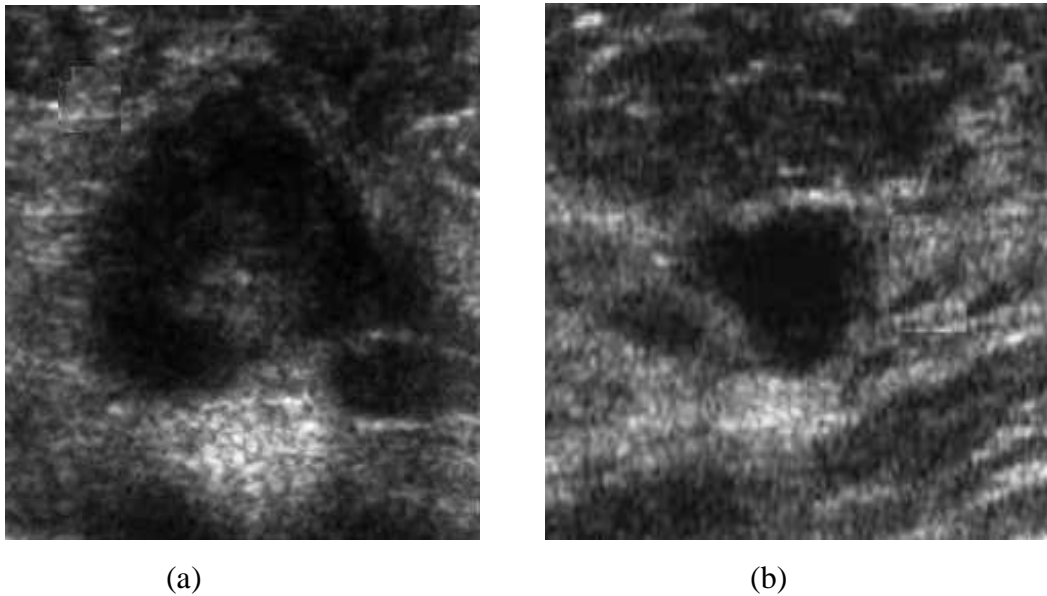


Figure 1.8 : Image échographique pour le sein (a) tumeur maligne et (b) tumeur bénigne (Cheikhrouhou, 2012)

Les techniques qui permettent un dépistage de masse doivent être suffisamment rentables et efficaces pour atteindre les masses. Une fois que le cancer du sein a été détecté par les tests de dépistage, des évaluations plus détaillées sont généralement effectuées à l'aide de modalités de diagnostic qui peuvent également être utilisées pour le diagnostic initial ; des modalités complémentaires sont utilisées pour donner aux médecins et aux cliniciens une plus grande confiance à leur diagnostic initial (Subbhuraam et al., 2014). Actuellement, les modalités utilisées comprennent la mammographie, l'échographie mammaire (voir figure 1.8), la thermographie, l'imagerie par résonance magnétique (IRM) (voir figure 1.9), la tomographie par émission de positrons, la scintimammographie, l'imagerie optique, l'imagerie basée sur l'impédance électrique et tomодensitométrie (voir figure 1.10) (Subbhuraam et al., 2014).

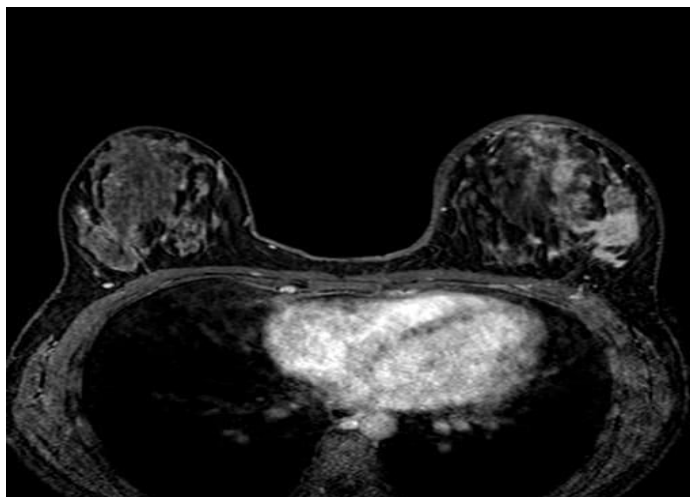


Figure 1.9 : IRM mammaire (web 03)

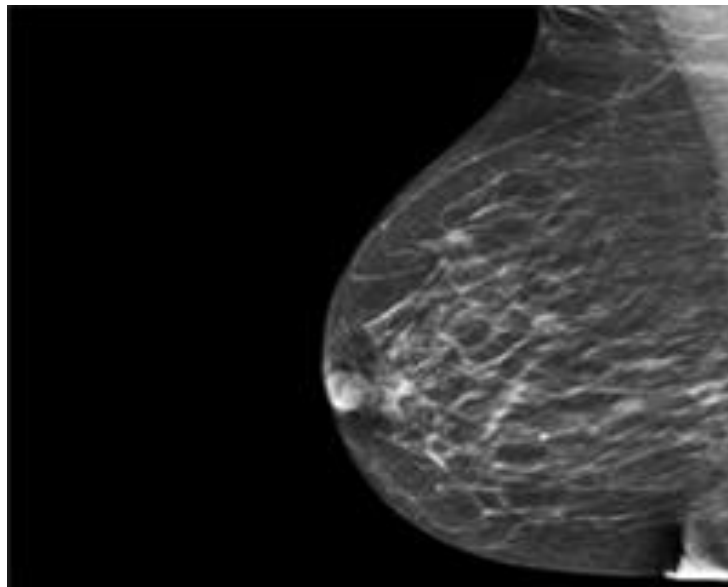


Figure 1.10 : Tomosynthèse mammaire (web 04)

Nelson et al. (Nelson et al., 2016) ont décrit quelques modalités d'image dédiées au dépistage du cancer du sein chez les femmes à risque moyen (elles présentant un risque de cancer du sein inférieur à 15 % au cours de leur vie) (voir tableau 1.4), où (BRCA : *breast cancer susceptibility gene* ; FDA : *U.S. Food and Drug Administration* ; mGy : *milligray* ; MRI : *magnetic resonance imaging* ; RCT : *randomized controlled trial*).

Description de la modalité d'imagerie ;	indication d'utilisation ; dose moyenne de rayonnement	Limitations
Mammographie	<ul style="list-style-type: none"> - Une mammographie de dépistage est effectuée chez une femme ne présentant aucun symptôme clinique afin de détecter un cancer du sein à un stade précoce. - Deux vues (crânio-caudale et oblique médio-latérale) de chaque sein sont obtenues pour une évaluation de routine. 	Les limitations varient selon le type de la mammographie
Mammographie sur film	<ul style="list-style-type: none"> - Utilise les rayons X transmis à travers le tissu mammaire pour créer une image qui est traitée et affichée en niveaux de gris directement sur un film. - Une compression adéquate des seins est nécessaire. - Les femmes ayant des seins plus volumineux peuvent avoir besoin de plus de deux clichés 	<ul style="list-style-type: none"> - Sensibilité limitée chez les femmes ayant des seins denses. - Incapacité de manipuler l'image après l'exposition.

	<p>de chaque sein pour garantir l'imagerie de tous les tissus mammaires.</p> <p>- La dose moyenne de rayonnement est de 4,7 mGy.</p>	
<p>Mammographie numérique DM (Digital mammography)</p>	<p>- Les détecteurs numériques convertissent les photons des rayons X en un signal électronique qui est transformé en une image numérique traitée et affichée sous forme d'image en niveau de gris afin d'être stocker ou envoyer par voie électronique. Un logiciel peut être utilisé pour aider à l'interprétions des images numériques.</p> <p>- Peut être plus efficace que la mammographie sur film chez les femmes <50 ans ; les femmes ayant un tissu mammaire hétérogène ou extrêmement dense, ou les femmes pré ou périménopausées.</p> <p>- La dose moyenne de rayonnement est de 3,7 mGy.</p>	<p>- Résolution spatiale inférieure à celle de la mammographie sur film.</p> <p>- Plus coûteuse (1,5 à 4 fois le coût de la mammographie sur film).</p>
<p>Tomosynthèse</p>	<p>- Une modification de la mammographie numérique qui permet d'acquérir des images d'un sein fixe et comprimé sous plusieurs angles au cours d'un court balayage.</p> <p>- Les images individuelles sont reconstruites pour générer une série de fines sections d'images qui peuvent être affichées individuellement ou en boucle.</p> <p>- Utilisé en combinaison avec la mammographie numérique standard pour le dépistage.</p> <p>- Dose moyenne de rayonnement de 1 à 2 fois la mammographie numérique.</p>	<p>Lorsqu'elle est effectuée dans le cadre du dépistage, la patiente est exposée à environ deux fois la dose de rayonnement habituelle, qui peut être encore plus importante si la patiente a des seins denses ou épais.</p>
<p>Échographie</p>	<p>- Ondes sonores utilisées pour créer des images du sein à l'aide d'un appareil manuel non invasif. Les images sont obtenues par un radiologue ou un technologue et dépendent de</p>	<p>- Ne constitue pas une modalité de dépistage initial appropriée pour le cancer du sein, mais elle a été approuvée</p>

	<p>l'opérateur. L'échographie du sein entier est récemment approuvée par la FDA pour le dépistage des patients ayant des seins denses.</p> <ul style="list-style-type: none"> - N'est pas actuellement indiqué pour le dépistage de routine. - Il n'existe pas de RCTs montrant un avantage en termes de survie du dépistage des femmes ayant des seins denses avec un dépistage supplémentaire par échographie du sein entier (échographie du sein entier) en plus de la mammographie. - Pas de radiation. 	<p>comme complément à la mammographie pour le dépistage chez les femmes ayant une densité mammaire élevée.</p> <ul style="list-style-type: none"> - L'échographie seule n'est pas un bon outil de dépistage du cancer du sein, elle donne de nombreux résultats faux positifs et faux négatifs. - Il n'existe pas de normes de performance uniformes. - Qualité d'image variable en fonction des compétences et de l'expérience de l'examineur. - Fortement dépendant de l'opérateur et il peut y avoir une grande variabilité intra- et inter-observateurs. - Capacité limitée à détecter le carcinome canalaire in situ.
IRM avec et sans contraste	<ul style="list-style-type: none"> - Les champs magnétiques sont utilisés pour créer une image du sein. Agent de contraste intraveineux donné pour l'intervention. - Non indiquée pour le dépistage dans les populations à risque moyen ; modalité de diagnostic dans des sous-populations spécifiques. - Pas de radiation. 	<p>Ne constitue pas une modalité de dépistage initiale appropriée pour le cancer du sein, mais elle a été promue comme test de dépistage chez les femmes à risque élevé, notamment les porteuses de la mutation BRCA1/2, les forts antécédents familiaux de cancer du sein ou plusieurs syndromes génétiques.</p>

Tableau 1.4 : Modalités d'imagerie pour le dépistage du cancer du sein chez les femmes à risque moyen (Nelson et al., 2016).

La mammographie est considérée comme la meilleure méthode de détection précoce du cancer du sein. Cependant, même si une mammographie mammaire peut montrer des régions suspectes, elle ne peut pas prouver qu'une zone anormale est cancéreuse ; Les vues mammographiques de dépistage les plus courantes sont les vues crânio-caudales (CC) et les vues obliques médio-latérales (MLO), la vue mammographique CC est saisie à partir du haut d'un sein comprimé horizontalement (le sein est comprimé à l'aide de plaques de compression), à son tour, le MLO est prise sur le côté et à l'angle d'un sein comprimé en diagonale (Abdel-Nasser et al., 2016).

La figure 1.11 illustre une image mammographique d'une femme de 48 ans présentant une masse palpable dans la partie supérieure du sein externe gauche qui montre un tissu hétérogène et dense qui peut masquer de petites masses ; le cancer de la patiente, dans la zone marquée par les flèches rouges, n'est pas bien vu.

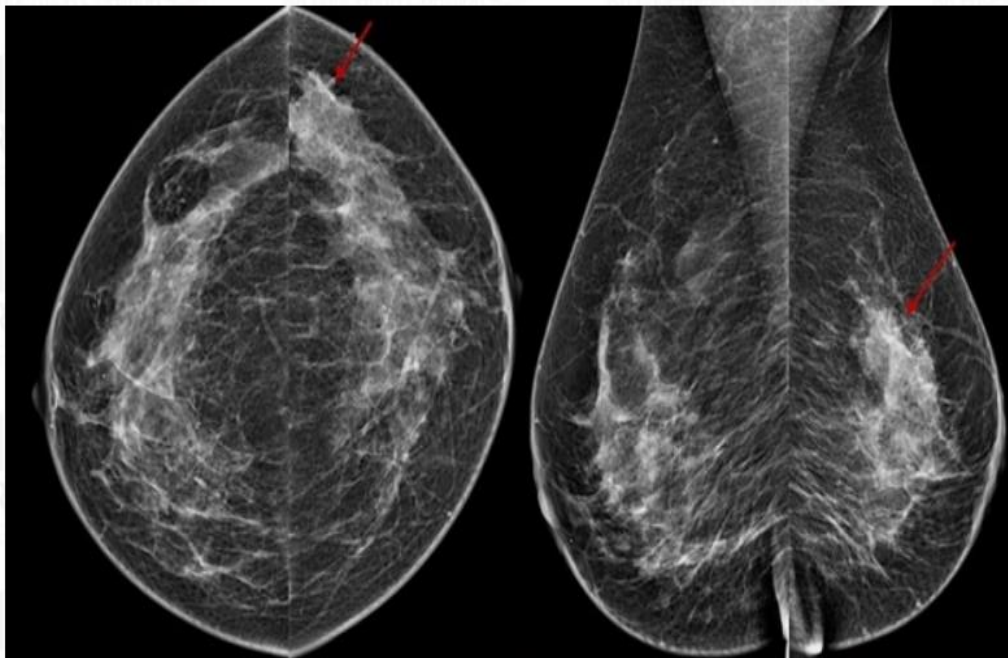


Figure 1.11 : Images standards d'une mammographie numérique : Les images crânio-caudales (CC, gauche) et médio-latérales obliques (MLO, droite) (web 05)

La mammographie ne peut pas détecter tous les types de cancer du sein, mais elle est néanmoins largement utilisée dans le monde entier pour la détection du cancer du sein en raison de son caractère peu coûteux et peu complexe ; elle permet de détecter environ 80 à 90 % des cancers du sein (Mussarat et al., 2013). La détection des masses ou d'anomalies à un stade précoce est tout à fait possible avec l'utilisation de la mammographie ; La zone du sein est extraite sous forme d'image et traitée avant d'être imprimée sur le film pour une meilleure

visualisation de la taille, de l'emplacement et de l'angle de la masse ; ces images optimisées, sont ensuite observées par le radiologue pour la détection d'éventuelles anomalies (Mussarat et al., 2013).

1.6 Traitements du cancer du sein

Le diagnostic et les traitements du cancer peuvent avoir un impact significatif sur la qualité de vie des patients (Grønvold, 2010). Les principaux traitements des cancers du sein sont la chirurgie, la radiothérapie, l'hormonothérapie et la chimiothérapie (Meunier, 2010).

1.6.1 La chirurgie :

L'intervention chirurgicale avec ou sans traitement adjuvant supplémentaire est un traitement loco-régional du cancer du sein. L'intervention chirurgicale consiste à enlever la tumeur du sein et les métastases dans les ganglions lymphatiques locaux, et implique soit une tumorectomie (également appelée lumpectomie), soit une mastectomie (ablation du sein). La chirurgie est la partie la plus importante du traitement du cancer du sein (Grønvold, 2010).

1.6.2 La radiothérapie

La radiothérapie consiste à utiliser des rayons X ou des rayons gamma de haute énergie, qui ciblent une tumeur ou le site d'une tumeur post-opératoire. Ces radiations sont très efficaces pour tuer les cellules cancéreuses qui peuvent rester après l'opération ou se reproduire là où la tumeur a été enlevée ; En plus de ce traitement, des cathéters radioactifs implantés (curiethérapie), similaires à ceux utilisés dans le traitement du cancer de la prostate, peuvent être utilisés, toutefois, cette option de traitement a été remplacée par la radiothérapie par faisceau d'électrons sur la cicatrice du sein. La radiothérapie pour le cancer du sein est généralement effectuée après une opération et elle fait partie intégrante de la thérapie de conservation du sein ; La dose de rayonnement doit être suffisamment forte pour assurer l'élimination des cellules cancéreuses. Les traitements sont généralement administrés sur une période de cinq à sept semaines (cinq jours par semaine), et chaque traitement dure environ 15 minutes (Sharma et al., 2010).

1.6.3 L'hormonothérapie

L'utilisation de l'hormonothérapie pour le cancer du sein est un exemple de l'une des premières utilisations de la thérapie ciblée pour le cancer (Puhalla et al., 2012). Elle est utilisée pour traiter les cancers du sein qui sont positifs aux récepteurs hormonaux, ces cancers ont des récepteurs pour les hormones œstrogène et/ou progestérone ; c'est pourquoi l'objectif principal de l'hormonothérapie est d'affamer les cellules du cancer du sein de

l'hormone (œstrogène) qui les fait croître, cela réduit le risque de réapparition du cancer du sein ou de développement d'un nouveau cancer du sein dans le sein traité ou dans l'autre sein (web 06). Une fois qu'une tumeur a été définie comme ayant une expression de récepteur d'œstrogène et/ou de progestérone (PR), il existe des stratégies potentielles pour cibler la voie hormonale (Puhalla et al., 2012), certaines sont prises sous forme de comprimés (tamoxifène ou inhibiteurs de l'aromatase) et d'autres peuvent impliquer une intervention chirurgicale, des injections ou une radiothérapie ; La thérapie hormonale peut être recommandée après d'autres traitements du cancer du sein comme la chirurgie, la chimiothérapie ou la radiothérapie, parfois, elle est utilisée pour réduire le cancer du sein avant que d'autres traitements ne soient administrés (web 06).

1.6.4 La chimiothérapie

La chimiothérapie est l'utilisation de médicaments anticancéreux pour traiter les cellules cancéreuses. Le traitement spécifique du cancer du sein sera basé sur l'état de santé général, les antécédents médicaux, l'âge (présence ou non de menstruations), le type et le stade du cancer, la tolérance à des médicaments et procédures spécifiques, etc. Les traitements de chimiothérapie sont souvent administrés par cycles ; un traitement pendant une certaine période, suivi d'une période de récupération, puis d'un autre traitement (Sharma et al., 2010). La chimiothérapie peut être administrée en préopératoire (néoadjuvant) pour réduire la tumeur et parfois permettre une chirurgie conservatrice du sein plutôt qu'une mastectomie, ou comme un traitement complémentaire après une chirurgie primaire (chimiothérapie adjuvante), qui peut être administré toutes les trois semaines ou toutes les deux semaines de manière "dense" (Sharma et al., 2010 ; Papakonstantinou, 2020).

1.7 Bases de données des images mammographiques numériques

1.7.1 La base DDSM

La base de données numérique pour le dépistage de la mammographie DDSM (*The Digital Database for Screening Mammography*) est une collection de mammographies provenant des sources suivantes : Hôpital général du Massachusetts, École de médecine de l'Université de Wake Forest, Hôpital du Sacré-Cœur et École de médecine de l'Université de Washington à St Louis (Lee et al., 2017).

La base de données DDSM contient environ 2 500 études, chaque étude comprend deux images de chaque sein, ainsi que certaines informations associées sur la patiente (âge au moment de l'étude, évaluation de la densité mammaire par ACR, évaluation de la subtilité des

anomalies, description des anomalies par mot-clé ACR) et des informations sur les images (scanner, résolution spatiale, etc.) ; Les images contenant des zones suspectes sont associées à des informations de "vérité de base" au niveau du pixel sur les emplacements et les types de régions suspectes (web 07).

1.7.2 La base AMDI

La base AMDI (*Indexed Atlas of Digital Mammograms*) est un système qui intègre des modules permettant l'ajout de nouveaux cas dans la base de données mammographiques par des radiologues autorisés, et qui aide les activités de recherche et d'éducation sur le cancer du sein grâce à une interface flexible et facile à utiliser via le Web (Guliatto et al., 2009).

Cette base de données mammographiques a été conçue pour inclure les cas avec toutes les vues mammographiques disponibles, les résultats radiologiques, le diagnostic prouvé par biopsie, l'historique clinique de la patiente et les informations concernant le style de vie de la patiente. Chaque examen de chaque cas comprend quatre vues CC ou MLO ; pour répondre aux aspects d'enseignement et de recherche, la base de données relie chaque mammographie au contour du sein, à la limite du muscle pectoral (vues MLO uniquement), aux contours des masses (si présentes), aux régions des groupes de calcifications et au nombre de calcifications (si présentes), et aux emplacements et détails de toute autre caractéristique d'intérêt ; La base de données AMDI permet également d'inclure plusieurs examens mammographiques de la même patiente effectués à des moments différents (web 08).

1.7.3 La base InBreast

La base de données InBreast a été acquise au Centre du sein du CHSJ, à Porto, Les images ont été acquises entre avril 2008 et juillet 2010 , elles ont été enregistrées au format DICOM (*Digital Imaging and Communications in Medicine*) qui rassemble non seulement l'image mais aussi certaines métadonnées connexes. Inbreast dispose d'images de mammographie numérique plein champ FFDM (*full field digital mammography*) provenant des cas de dépistage, de diagnostic et de suivi, où le dépistage est effectué conformément aux normes nationales et régionales (Moreira et al., 2012). Le diagnostic est effectué lorsque le dépistage montre des signes d'anomalie, et dans les images de suivi, le cancer a été préalablement détecté et traité. Au total, 115 cas ont été recueillis, dont 90 ont deux images (MLO et CC) de chaque sein et les 25 cas restants proviennent des femmes ayant subi une mastectomie ; deux vues d'un seul sein ont été incluses. La base de données InBreast comprend des exemples de mammographies normales, de mammographies avec masses, de mammographies avec calcifications, de déformations architecturales (Moreira et al., 2012)

1.7.4 La base MAGIC-5

MAGIC-5 est une base de données mammographique italienne d'images numérisées pour a recherche (en italien : *un database mammografico italiano di immagini digitalizzate per scopi di ricerca*) est composée de 3369 images mammographiques, chacune contient des données et des informations cliniques. Les images ont été recueillies auprès de 967 patients. Les groupes d'âge sont indiqués dans la figure 1.12, où environ 60 % des patients de cette base de données ont plus de 50 ans (Tangaro et al., 2008). Toutes les images mammographiques ont des informations supplémentaires liées au patiente (suivi, âge, et intérêt), elles contiennent une (ou plusieurs) lésion ; elles ont été classées selon le type de lésion (masse ou microcalcification), le degré de malignité, le type de texture des seins, etc (Tangaro et al., 2008).

MAGIC-5 contient les images de 306 (32%) patients qui ont été définies comme normaux. Les autres images proviennent de 661 (68%) anormales. Les annotations radiologiques qui détaillent des anomalies sont incluses dans la base de données sous forme de notes. La distribution relative du grade de malignité est de 560 (35%) lésions suspectes, 468 (29%) lésions bénignes et 592 (37%) lésions malignes (Tangaro et al., 2008).

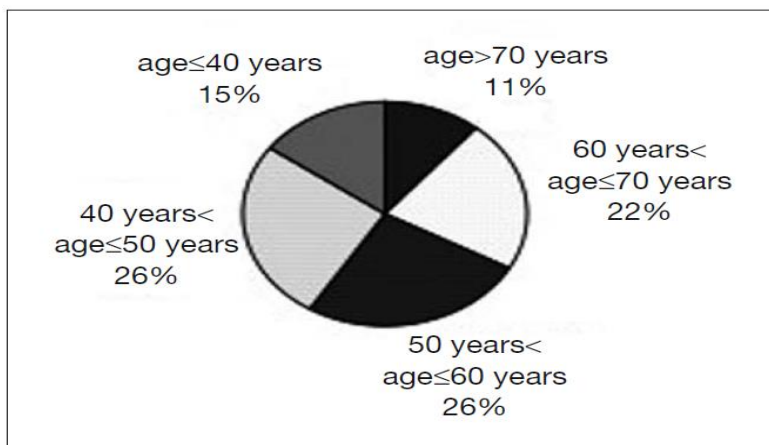


Figure 1.12 : Groupes d'âge des patientes (Tangaro et al., 2008).

1.7.5 La base WDBC

La base WDBC (*Wisconsin Diagnosis Breast Cancer Database*) est une base qui est distribuée par Dr William HWolberg et autres (Wolberg and Mangasarian, 1990) ; Cette base est constituée de données provenant de 569 cas (357 sont bénins et 212 sont malins) (Daoudi and Khalifa, 2016).

Certaines des caractéristiques incluses dans l'ensemble de données de la base WDBC sont des propriétés telles que le rayon, la texture, le périmètre, la surface, la douceur, la compacité, la

concavité, les points concaves, la symétrie, la taille fractale pour chaque noyau cellulaire ; de nombreuses études ont été réalisées sur l'ensemble des données relatives au cancer du sein de la base WDBC et ils ont montré un grand succès (Saygılı, 2018).

1.8 Conclusion

Au cours de ce chapitre, nous avons exposé le fond théorique médical de la maladie du cancer du sein.

Dans un premier temps, nous avons abordé le développement du cancer du sein dans le monde, nous avons également présenté plusieurs notions liées à cette maladie : l'incidence, la mortalité, l'anatomie de la poitrine féminine, les étiologies, le dépistage, les modalités d'imagerie médicale qui permettent son bon diagnostic, ainsi que son traitement.

Dans un deuxième temps, nous avons présenté quelques bases de données des images mammographiques, telles que DDSM, AMDI, InBreast, MAGIC-5, et WDBC.

Les concepts généraux sur la classification des images seront présentés dans le prochain chapitre.

Chapitre 02

Concepts de base sur la classification des images

2.1 Introduction

La classification automatique des images est un problème très courant, elle couvre plusieurs domaines (la biométrie, la médecine, l'industrie...etc.)

Dans le domaine médical, l'utilisation des systèmes automatique d'aide au diagnostic est très avantageuse ; les étapes primordiales dans le processus complet de ces systèmes sont les étapes « extraction des descripteurs » et « classification ».

Le présent chapitre est consacré à la description du fond théorique des notions de base utilisées tout au long de notre travail. Dans un premier temps, nous présentons le principe général d'un système d'aide au diagnostic (SAD) ainsi que ses différents composants. Dans un deuxième temps, nous exposons les notions théoriques des descripteurs de texture et des méthodes de classification des images ; Des concepts généraux sur la technique de la super-résolution seront abordés dans le reste de ce chapitre.

2.2 Système de diagnostic assisté par ordinateur

Généralement, le système de diagnostic assisté par ordinateur qui a pour but l'analyse des images mammographiques est un processus complet qui doit se dérouler consécutivement, Commençons par l'acquisition des images jusqu'à la l'étape finale de classification, certaines de ces phases sont vraiment fortement liées et indispensable. Les étapes essentielles de traitement d'une image mammographique sont (Cheikhrouhou, 2012) :

- L'étape de prétraitement.
- L'étape de segmentation.
- L'étape de description.
- L'étape de classification.

La plupart des algorithmes de traitement des images mammographiques se composent des étapes précédentes (voir figure 2.1) ; tout d'abord, les images mammographiques sur le film d'écran doivent être numérisées avant le traitement, c'est l'une des avancées de la mammographie numérique où l'image peut être traitée directement (Bozek et al., 2009).

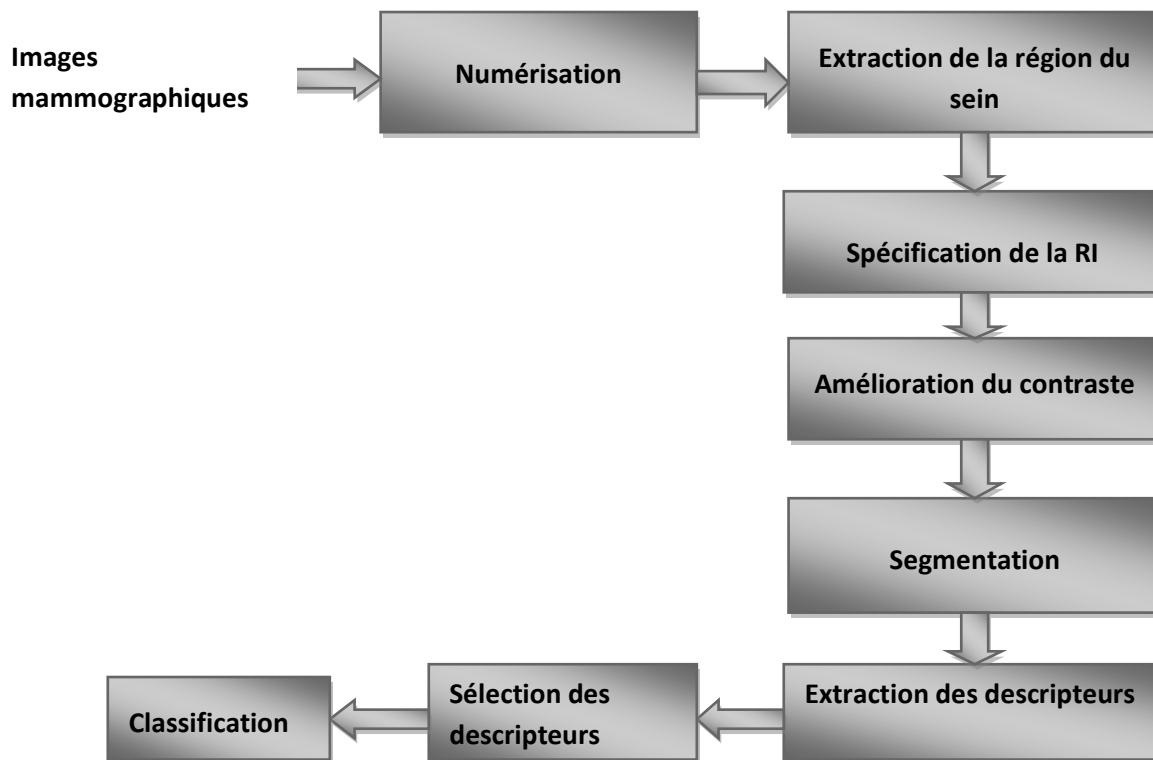


Figure 2.1 : Processus général d'un SAD dédié à l'image mammographique (Mabrouk et al., 2019)

L'étape suivante est le prétraitement, qui doit être réalisé sur des images numérisées pour améliorer la finesse de l'image originale, et pour augmenter la différence d'intensité entre les objets et le fond (Mabrouk et al., 2019); la plupart des images mammographiques numériques sont des images de haute qualité, une autre partie de l'étape de prétraitement consiste à supprimer la zone du fond et l'ablation du muscle pectoral de la zone du sein si l'image est une vue MLO (Bozek et al., 2009).

L'étape de segmentation vise à trouver des régions d'intérêt (ROI) suspectes contenant des anomalies ; Dans l'étape d'extraction des caractéristiques, les caractéristiques sont calculées à partir des caractéristiques de la région d'intérêt (Bozek et al., 2009). La question essentielle dans la conception dans un système SAD est comment choisir le meilleur ensemble de caractéristiques afin d'éliminer les faux positifs ?, la sélection de ces caractéristiques est définie comme la sélection d'un sous-ensemble de caractéristiques plus petit qui conduit à la

bonne classification (Bozek et al., 2009), qui est la dernière étape du système SAD où le classifieur utilise les caractéristiques importantes et robustes acquises lors de l'étape de sélection des caractéristiques pour classer la lésion anormale en bénigne ou maligne (Mabrouk et al., 2019).

L'extraction des caractéristiques et la classification sont les phases clés dans un système d'aide au diagnostic, c'est pourquoi nous exposons le fond théorique de ces deux phases dans la prochaine section.

2.3 Concepts et méthodes de base pour la classification des images

Ces dernières années, et avec le développement de la technologie de traitement des images numériques et de l'intelligence artificielle, l'utilisation d'un système de détection et de diagnostic assisté par ordinateur est devenue un problème scientifique réaliste et important ; De nombreux chercheurs ont également utilisé le système SAD pour identifier et classer les micro calcifications et les masses dans les images mammographiques (Li et al., 2019).

Le but des systèmes de détection assistés par ordinateur est d'aider les radiologues à éviter d'oublier les caractéristiques anormales qui ne sont pas visibles sur l'image de dépistage (Jalalian et al., 2013) et ils les aident à juger la classification des tumeurs du sein (Li et al., 2019).

2.3.1 Extraction des caractéristiques

L'extraction de caractéristiques est le processus qui consiste à définir un ensemble de caractéristiques qui représentent les informations importantes pour l'analyse et la classification (Saraswathi and Srinivasan, 2016).

Dans un système SAD, l'extraction des caractéristiques occupe une place très importante car son efficacité globale dépend énormément sur les caractéristiques discriminantes utilisées (Rabidas et al., 2016). Dans les étapes d'extraction et de sélection des caractéristiques, les caractéristiques qui caractérisent les régions d'intérêt sont calculées et celles qui sont importantes sont sélectionnées pour la classification des masses (maligne/bénigne) (Bozek et al., 2009).

L'espace de présentation de caractéristique est très vaste et complexe en raison de la grande diversité des tissus normaux et de la variété des anomalies, certaines des caractéristiques ne sont pas significatives lorsqu'elles sont observées seules, mais la combinaison avec d'autres caractéristiques peut être importante pour la classification (Bozek et al., 2009).

Li et al. (Li et al., 2001) a proposé des lignes directrices générales pour l'extraction et la sélection des caractéristiques importantes : la discrimination, la fiabilité, l'indépendance et l'optimalité, et selon la littérature de El Atlas et al. et Martinez (El Atlas et al., 2014; Martinez, 2004), les descripteurs d'images en général, sont classés en trois catégories, à savoir les descripteurs basés sur la texture, la couleur, et la forme qui sont des propriétés importantes utilisées pour distinguer les objets (Jalalian et al., 2017). Les trois descripteurs sont résumés dans la figure 2.2.

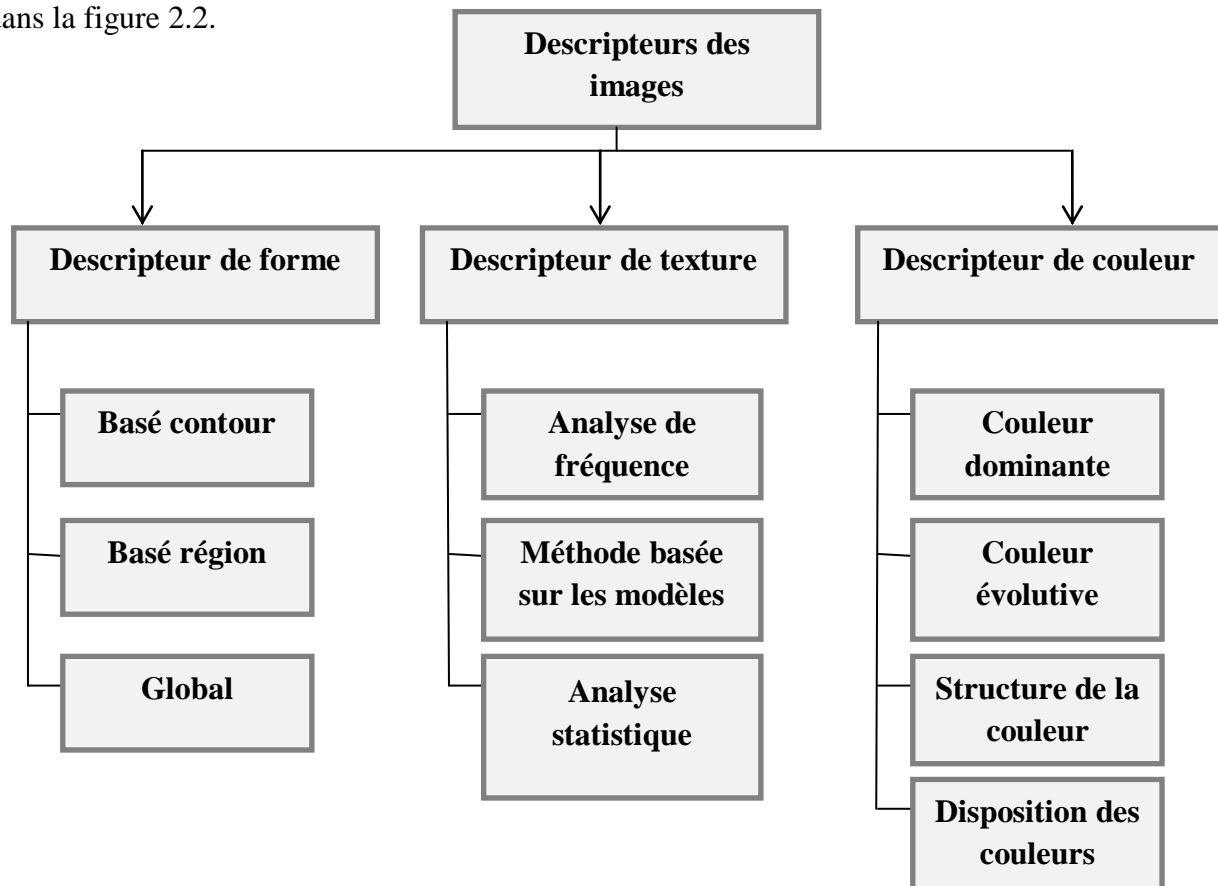


Figure 2.2 : Les descripteurs des images (Jalalian et al., 2017).

Pour traiter la conversion de la complexité des formes, un descripteur de forme efficace doit être invariant dans la rotation et la mise à l'échelle ; les descripteurs basés sur la couleur sont des repères visuels importants pour la recherche d'images et la reconnaissance d'objets ; les descripteurs de texture sont les caractéristiques les plus courantes utilisées pour analyser et interpréter les images en considérant la variation d'intensité et en quantifiant différentes propriétés telles que la douceur, la grossièreté et la régularité (Jalalian et al., 2017).

Dans le cadre de cette thèse nous sommes intéressés à l'étape d'extraction des caractéristiques de texture ; c'est pourquoi nous exposons les principaux descripteurs de texture dans cette section.

2.3.1.1 La notion de texture

Il existe plusieurs définitions de la texture :

- La texture est une propriété qui représente la surface et la structure d'une image, d'une manière générale, la texture peut être définie comme la répétition régulière d'un élément ou d'un motif sur une surface (Srinivasan, and Shobha, 2008).
- Les textures d'images sont des motifs visuels complexes composés des entités ou des régions avec des sous-composantes ayant les caractéristiques suivantes : la luminosité, la couleur, la forme, la taille, etc ; Une région d'image a une texture constante si un ensemble de ses caractéristiques sont constantes, changeant lentement ou approximativement de façon périodique ((Sklansky, 1978 dans Srinivasan, and (Shobha, 2008)).
- La texture peut être considérée comme un regroupement de similitudes dans une image ((Rosenfeld, 1976) dans (Srinivasan, and Shobha, 2008)).
- La texture d'une région de l'image est déterminée par la façon dont les niveaux de gris sont répartis sur les pixels de la région, bien qu'il n'y ait pas de définition claire de la "texture" dans la littérature, elle décrit souvent l'aspect d'une image par une apparence fine ou grossière, lisse ou irrégulière, homogène ou inhomogène,...etc. (Aggarwal and Agarwal, 2012).



Figure 2.3 : Exemple des images ont une texture naturelle (Suresha, 2017).

Dans les images naturelles, la texture se présente sous la forme d'une mosaïque de différents types de textures ; la figure 2.3 illustre des exemples des images ont une texture naturelle ; la

figure 2.4 montre un exemple des images ont une texture régulière artificielle ; la figure 2.5 montre un exemple des images ont une texture régulière naturelle (Suresha, 2017).



Figure 2.4 : Exemple des images ont une texture régulière artificielle (Suresha, 2017).

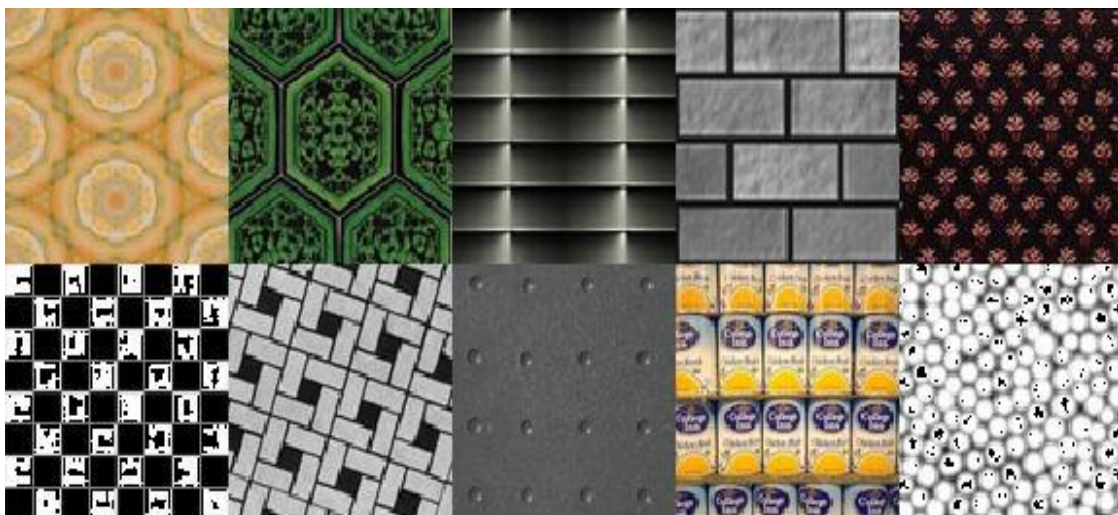


Figure 2.5 : Exemple des images ont une texture régulière naturelle (Suresha, 2017).

Les caractéristiques de texture sont décrites pour quantifier les propriétés d'une région de l'image en exploitant les relations spatiales sous-jacentes à la distribution des niveaux de gris d'une image donnée (Aggarwal and Agarwal, 2012), la texture dans l'image peut être extraite à l'aide des outils mathématiques permettant d'obtenir des descripteurs de texture (Attia, 2013). Dans la littérature il existe quatre classes d'attributs de texture :

- Les descripteurs géométriques : qui dépends de l'information structurelle et contextuelle d l'image ; l'extraction des primitives (contours, régions, ...) et de règles de placement de ces primitives décrit ce type de texture (Porebski, 2009).

- Les descripteurs basés sur la modélisation spatiale des textures : tient compte de la répartition spatiale des couleurs dans l'image, qui est modalisée par (les modèles autorégressifs multi spectraux, les modèles de MARKOV et les motifs locaux binaires) (Porebski, 2009).
- Les descripteurs spatio-fréquentiels : se calculent dans le domaine spatial, dans le domaine fréquentiel ou dans le domaine spatio-fréquentiel (la transformée de Gabor, la transformée d'ondelette, et la transformée de fourrier) (Attia, 2013).
- Les descripteurs statistiques : qui traduisent la relation entre un pixel et ses voisins. Il existe trois types d'attributs statistiques (premier ordre, deuxième ordre, et ordre supérieur) (Attia, 2013).

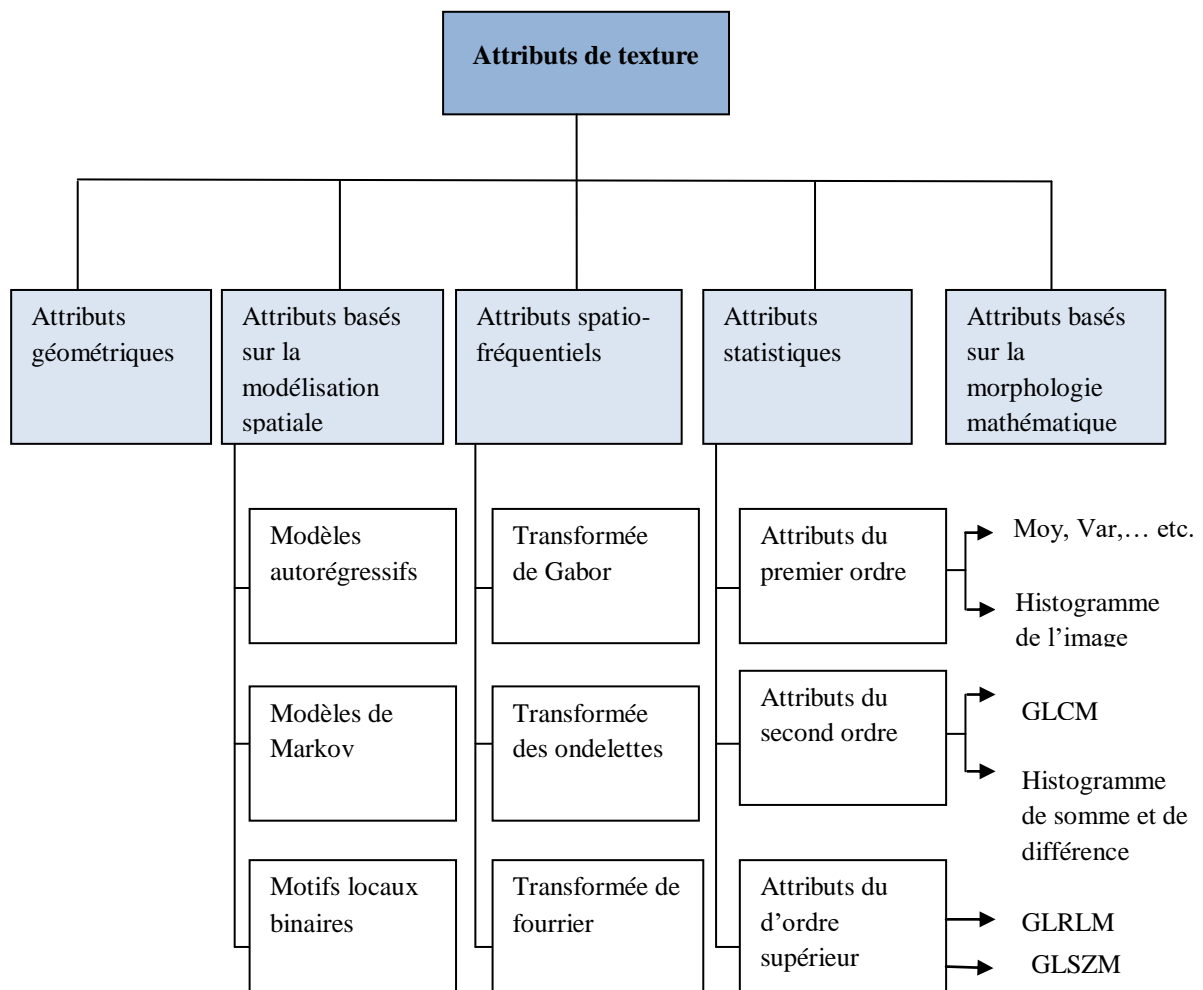


Figure 2.6 : Les différentes classes d'attributs de texture de la littérature (Attia, 2013).

Attia (Attia, 2013) a considéré une cinquième catégorie de descripteurs de texture qui est calculée à l'aide des opérateurs de la morphologie mathématique qui s'appuient sur la théorie

des ensembles, les treillis, la topologie des ensembles fermés et les probabilités. La figure 2.6 résume les cinq classes d'attributs de texture.

Dans le cadre de cette thèse, notre attention se porte sur les descripteurs statistiques et les motifs binaires locaux, c'est pourquoi, nous les décrivons dans le reste de cette section.

2.3.1.2 Les attributs statistiques

Les méthodes statistiques analysent la distribution spatiale du niveau de gris en calculant les caractéristiques locales à chaque point de l'image, et en dérivant un ensemble de statistiques à partir des distributions des caractéristiques locales (Srinivasan and Shobha, 2008).

La distribution spatiale des valeurs de niveau de gris est l'une des caractéristiques importante de la texture ; en fonction du nombre de pixels qui définit les méthodes statistiques, elles peuvent être classées en premier ordre (un pixel), second ordre (deux pixels) et d'ordre supérieur (trois pixels ou plus) (Ojala and Pietikäinen, 2004 ; Srinivasan and Shobha, 2008); La différence fondamentale est que les statistiques du premier ordre permettent d'estimer les propriétés (par exemple, les variances) des valeurs de chaque pixel, en ignorant l'interaction entre les pixels de l'image, alors que les statistiques du second ordre et du rang supérieur estiment les propriétés de deux ou plusieurs valeurs de pixels qui se produisant à des endroits spécifiques les uns par rapport aux autres (Srinivasan and Shobha, 2008).

➤ Attributs statistiques du premier ordre

Les Attributs statistiques de texture du premier ordre sont calculées à partir des valeurs de l'image originale, comme la variance, et ils ne tiennent pas compte les relations de voisinage des pixels (Srinivasan and Shobha, 2008). L'approche statistique d'analyse de texture est basée sur l'histogramme qui est un graphique qui montre le nombre de pixels d'une image à chaque intensité différente trouvée dans l'image (Nithya and Santhi, 2011).

Le principal avantage de cette approche réside dans sa simplicité d'utilisation des mesures classiques, afin de caractériser des données (Alsheh Ali, 2015) ; les caractéristiques du premier ordre les plus utilisées sont : la moyenne (Sachdeva et al., 2013), la variance (Sachdeva et al. 2013), l'entropie (Nurhayati et al., 2011), l'asymétrie (Doane et Lori 2011), l'uniformité (Nithya and Santhi, 2007), régularité (Padma and Vijayalakshmi, 2016) et l'aplatissement (Sachdeva et al. 2013) , les formules de ces descripteurs sont présentés ci-dessous ; où la moyenne est la mesure de dispersion de l'image concernant la distribution des couleurs d'une image ; La variance exprime une distribution largement répandue. L'asymétrie

exprime une mesure d'asymétrie ; La distribution est dite de gauche si elle a une valeur d'asymétrie sous la forme d'un nombre négatif. Inversement, la distribution est dite penchée vers la droite si elle a une valeur d'asymétrie sous la forme d'un nombre positif. L'aplatissement est une mesure qui montre la distribution des données qui sont effilées ou collectées (Liantoni et al., 2019).

$$\text{Moyenne (m)} = \frac{1}{N} \sum_i \sum_j P(i, j) \quad (1)$$

$$\text{Variance (std)} = \sqrt{\frac{\sum_i \sum_j (P(i, j) - m)^2}{N}} \quad (2)$$

$$\text{Asymétrie} = \frac{\sum_{i=0}^{n-1} P_i (x_i - m)^3}{\sum_{i=0}^{n-1} P_i (x_i - m)^{3/2}} \quad (3)$$

$$\text{Aplatissement} = \frac{1}{N} \frac{\sum_i \sum_j (P(i, j) - m)^4}{\text{Std}^4} \quad (4)$$

$$\text{Entropie} = - \sum_i \sum_{ij} P(i, j) \log P(i, j) \quad (5)$$

$$\text{Uniformité} = \sum_i P_i^2 (z_i) \quad (6)$$

$$\text{Régularité} = 1 - \frac{1}{1 + \sigma^2(z)} \quad (7)$$

Où : $\mathbf{P(i, j)}$ est l'intensité du pixel à la position $(\mathbf{i, j})$

\mathbf{N} représente le nombre total des pixels

$\sigma(\mathbf{z})$ est l'écart-type du valeur du niveau de gris \mathbf{z} .

➤ Attributs statistiques du second ordre

La matrice de cooccurrences est parmi les approches les plus connues et les plus utilisées pour extraire les caractéristiques de texture du second ordre (Thibault, 2009). Les matrices de cooccurrences (GLCM), introduites par Haralick en 1973, ont tout d'abord été implémentées en niveaux de gris (Haralick et al., 1973). Cette méthode, malgré très ancienne, reste toujours très utile dans de beaucoup applications où l'information de texture est vraiment importante, notamment médicales (Garnier, 2014). Les caractéristiques de second ordre de la texture sont extraites à partir du matrice GLCM, qui est basée sur les pixels voisins (Nithya and Santhi, 2011), parfois elle est appelée la matrice de dépendance spatiale des tons de gris.

L'élément de matrice $\mathbf{P}(\mathbf{i}, \mathbf{j} | \mathbf{d},)$ contient les valeurs de probabilité statistique du second ordre pour les changements entre les niveaux de gris \mathbf{i} et \mathbf{j} à une distance de déplacement \mathbf{d} particulière et à un angle particulier θ ($0^\circ, 45^\circ, 90^\circ, 180^\circ$) (Alasadi and Baiee, 2013) (voir figure 2.7).

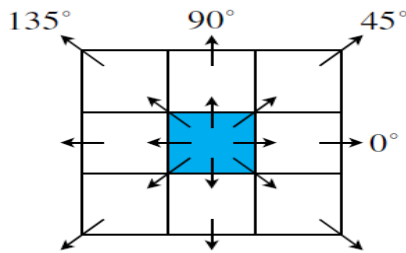


Figure 2.7 : Les quatre directions considérées (Porebski, 2009).

Au lieu d'utiliser directement les valeurs de fréquence dans un GLCM, il est courant de les normaliser dans la plage $[0, 1]$ afin d'éviter les effets d'échelle (Alasadi and Baiee, 2013). La figure 2.8 illustre un exemple de construction d'une matrice de co-occurrence où l'image est de taille $3 * 3$ et sa matrice GLCM cooccurrence de 4 niveaux de gris. Le nombre de niveaux seuils est 4, le 2 dans la matrice de cooccurrence indique qu'il y a deux occurrences d'un pixel avec le niveau de gris 3 immédiatement à droite du pixel avec niveau de gris 1 (Srinivasan and Shobha, 2008).

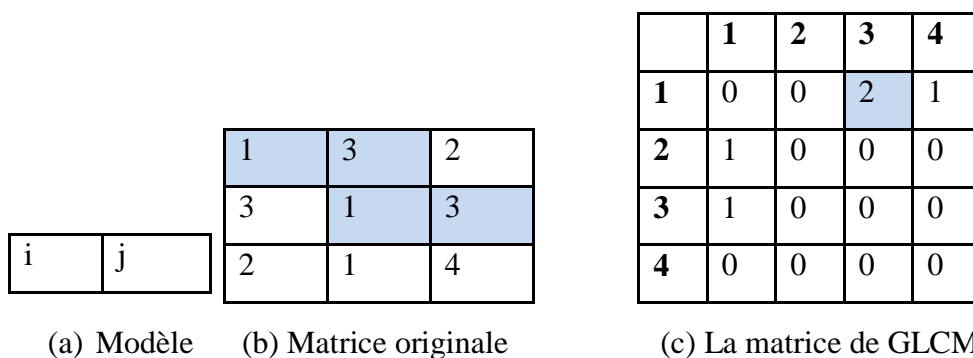


Figure 2.8 : Exemple de construction de la matrice GLCM (Srinivasan and Shobha, 2008).

Haralick et al. (Haralick et al., 1973) ont proposé quatorze attributs de texture extraits des matrices de co-occurrences, et parmi eux il y a six caractéristiques qui sont considérées comme étant les plus utilisées et les plus pertinentes (l'énergie, le contraste, l'entropie, la variance, la corrélation et le moment des différences inverses).

$$\text{Energie} = \sum_i \sum_j P_i(i, j)^2 \quad (8)$$

$$\text{Contraste} = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P(i, j) \right. \\ \left. |i - j| = n \right\} \quad (9)$$

$$\text{Corrélation} = \frac{\sum_i \sum_j (ij) P_i(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (10)$$

Où $\mu_x, \mu_y, \sigma_x,$ et σ_y sont les moyennes et les écarts type de P_x et P_y .

$$\text{Variance} = \sum_i \sum_j (i - \mu)^2 P(i, j) \quad (11)$$

$$\text{Moment des différences inverses} = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i, j) \quad (12)$$

$$\text{Moyenne des sommes} = \sum_{i=2}^{2Ng} i P_x + y (i) \quad (13)$$

$$\text{Variance des sommes} = \sum_{i=2}^{2Ng} (i - \text{entropie des sommes})^2 P_x + y (i) \quad (14)$$

$$\text{Entropie des sommes} = - \sum_{i=2}^{2Ng} P_x + y (i) \log \{P_x + y (i)\} \quad (15)$$

$$\text{Entropie} = - \sum_i \sum_j P(i, j) \log (P(i, j)) \quad (16)$$

$$\text{Variance des différences} = \text{variance} (P_x - y) \quad (17)$$

$$\text{Entropie des différences} = - \sum_{n=0}^{Ng-1} P_x - y (i) \log \{P_x - y (i)\} \quad (18)$$

$$\text{Information sur la corrélation 1} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (19)$$

$$\text{Information sur la corrélation 2} = (1 - \exp [-2.0(HXY2 - HXY)])^{1/2} \quad (20)$$

$$HXY = - \sum_i \sum_j P(i, j) \log (P(i, j))$$

Où :

HX et HY sont les entropies de P_x et P_y

$$HXY1 = - \sum_i \sum_j P(i, j) \log \{P_x(i) P_y(j)\}$$

$$HXY2 = - \sum_i \sum_j P_x(i) P_y(j) \log \{P_x(i) P_y(j)\}$$

Coefficient de corrélation maximal = (La grande valeur propre de Q)^{1/2}

Où :

$$Q(i,j) = \sum_k \frac{P(i,k)P(j,k)}{Px(i)Py(k)}$$

$P(i,j)$ c'est l'intersection des coordonnées (i,j) de la matrice.

N_g c'est le nombre de niveau de gris

$P_x(i)$ est la somme des éléments de la colonne i de la matrice.

$P_y(j)$ est la somme des éléments de la ligne j de la matrice.

$P_{x+y}(i)$ est la somme des éléments de la diagonale principale i de la matrice.

$P_{x-y}(i)$ est la différence des éléments de la diagonale secondaire i de la matrice.

Plusieurs autres auteurs ont également proposé d'autres attributs de cooccurrences comme Anys (Porebski, 2009), nous citons :

$$\text{Moyenne} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} i * P(i,j) \quad (21)$$

$$\text{Cluster shade} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i + j - ux - uy)^3 P(i,j) \quad (22)$$

$$\text{Cluster Prominence} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i + j - ux - uy)^4 P(i,j) \quad (23)$$

$$\text{Dissimilarité} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} |i - j| * P(i,j) \quad (24)$$

$$\text{Covariance} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i - \mu_x)(j - \mu_y) * P(i,j) \quad (25)$$

$$\text{Moment diagonal} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \left(\frac{1}{2}|i - j|P(i,j)\right)^{1/2} \quad (26)$$

$$\text{Différence inverse} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{P(i,j)}{1+|i-j|} \quad (27)$$

$$\text{Troisième moment angulaire} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (p(i,j))^3 \quad (28)$$

$$\text{Probabilité maximale} = \max_{0 \leq i, j \leq N_g-1} \{ p(i,j) \} \quad (29)$$

Les histogrammes des sommes et des différences ont été également considérés comme des attributs statistiques d'ordre deux, ils sont proposés par Unser (Unser, 1986) afin de caractériser les textures dans des images en niveaux de gris (Porebski, 2009) ; Les histogrammes des sommes et des différences ont une capacité de discrimination presque similaire à la GLCM, avec l'avantage de calculs qui est moins exigeants en espace mémoire (Porebski, 2009). La liste des caractéristiques de ces histogrammes se trouve en **annexe A**.

➤ Attributs statistiques d'ordre supérieur (Matrices de longueurs de plages)

La méthode des matrices de longueurs de plages GLRLM (*Gray Level Run Length Matrix*) proposée par Galloway (Galloway, 1975), analyse les images en considérant les segments de pixels qui ont le même niveau de composante (couleur ou niveau de gris), dans une direction ' θ ' (Zheng et al., 2007 dans Porebski, 2009), la longueur de plage est une suite de pixels successifs et adjacents qui ont tous le même niveau de composante (Porebski, 2009).

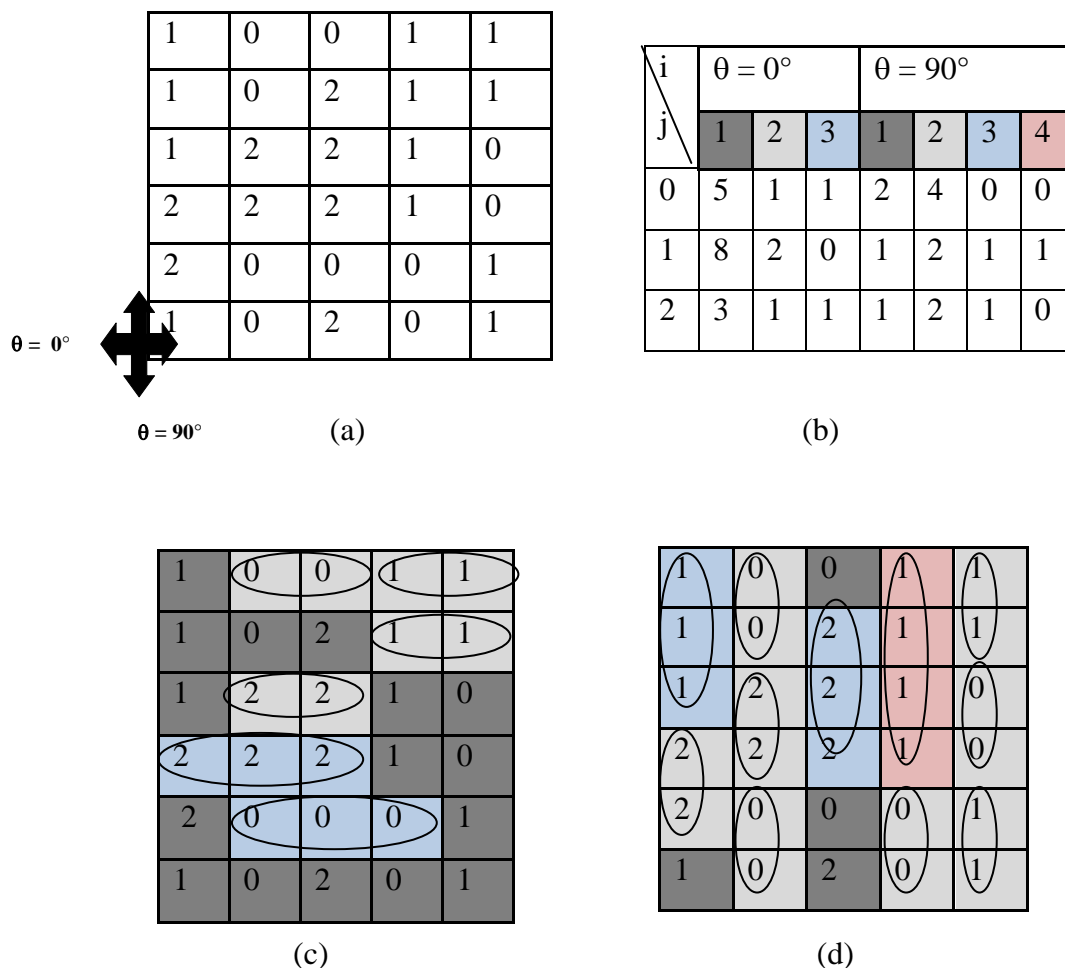


Figure 2.9 : Exemple de calcul des matrices GLRLM.

Le calcul de la longueur des segments est symétrique c.-à-d. que la longueur d'un segment est la même dans les directions (θ et $\theta + \pi$) et il est donc inutile de calculer la matrice dans les quatre directions opposées (180° , 225° , 270° ou 315°) ; Cette méthode nécessite moins de calculs que les matrices GLCMs (Thibault, 2009). La figure 2.9 montre un exemple de construction de deux matrices de longueurs de plages à quatre niveaux de gris dans les directions 0° et 90° .

Ci-dessous les caractéristiques d'ordre supérieur extraites à partir de la matrice de longueurs de plages (Filipczuk et al., 2012) sont présentées :

$$\text{Short Run Emphasis (SRE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j)}{j^2} \quad (30)$$

$$\text{Long Run Emphasis (LRE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P(i,j) j^2 \quad (31)$$

$$\text{Low Gray Level Run Emphasis (LGLRE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j)}{i^2} \quad (32)$$

$$\text{High Gray Level Run Emphasis (HGLRE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P(i,j) i^2 \quad (33)$$

$$\text{Short Run Low Gray Level Emphasis (SRLGLE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j)}{i^2 j^2} \quad (34)$$

$$\text{Short Run High Gray Level Emphasis (SRHGLE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) i^2}{j^2} \quad (35)$$

$$\text{Long Run Low Gray Level Emphasis (LRLGLE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) j^2}{i^2} \quad (36)$$

$$\text{Long Run High Gray Level Emphasis (LRHGLE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P(i,j) i^2 j^2 \quad (37)$$

$$\text{Gray Level Non Uniformity (GLN)} = \frac{1}{n_r} \sum_{i=1}^M \left(\sum_{j=1}^N P(i,j) \right)^2 \quad (38)$$

$$\text{Run Length Non Uniformity (RLN)} = \frac{1}{n_r} \sum_{j=1}^N \left(\sum_{i=1}^M P(i,j) \right)^2 \quad (39)$$

$$\text{Run Percentage (RP)} = \frac{n_r}{n_p} \quad (40)$$

Où :

- $\mathbf{P(i, j)}$ l'élément de coordonnées (i, j) de la matrice.
- n_r le nombre de segments de la matrice, $n_r = \sum_{i,j} P(i, j)$

Les attributs de la matrice de Gray Levels Size Zone Matrix (GLSZM) ont été également considérés comme des attributs statistiques d'ordre supérieur. Thibault (Thibault, 2009) a élaboré la matrice GLSZM qui représente le dénombrement des tailles de toutes les régions de texels de même niveau d'intensité, sa construction est similaire à la matrice GLRLM. La liste des caractéristique extraites de la matrice GLSZM se trouve en **annexe A**.

2.3.1.3 Attribut basé sur la modélisation spatiale des textures (méthode des motifs locaux binaires)

Les attributs basés sur la modélisation spatiale des textures sont issus de différentes modélisations de la répartition spatiale des couleurs dans l'image (Porebski, 2009), la méthode des motifs binaires est parmi les modèles les plus utilisés pour décrire la texture.

La méthode des motifs locaux binaires (*Local Binary Patterns* LBP), est introduite par (Ojala et al., 1996), elle est trop utile dans plusieurs domaines comme la reconnaissance faciale, et en terme d'efficacité discriminante, elle est performante et englobe des informations structurelles et statistiques (Paulhac, 2009).

Afin de calculer 'un code LBP' dans un voisinage de '**P**' pixels, dans un rayon '**r**', il faut compter les occurrences de niveaux de gris '**gp**' plus grands ou égaux à la valeur centrale (Houam, 2013), l'opérateur LBP donc étiquette le pixel d'une image en fixant un seuil au voisinage de chaque pixel avec la valeur centrale et le résultat c'est un nombre binaire, où, **s(x)** est la fonction de seuillage, **gc** est la valeur de niveau de gris du pixel central, **gp** est la valeur de niveau de gris de ses voisins (voir figure 2.10), **p** est le nombre de voisins impliqués (Sabu et al., 2012).

$$LBP_{p,r} = \sum_{p=0}^{p-1} s(gp - gc) 2^p \quad (41)$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (42)$$

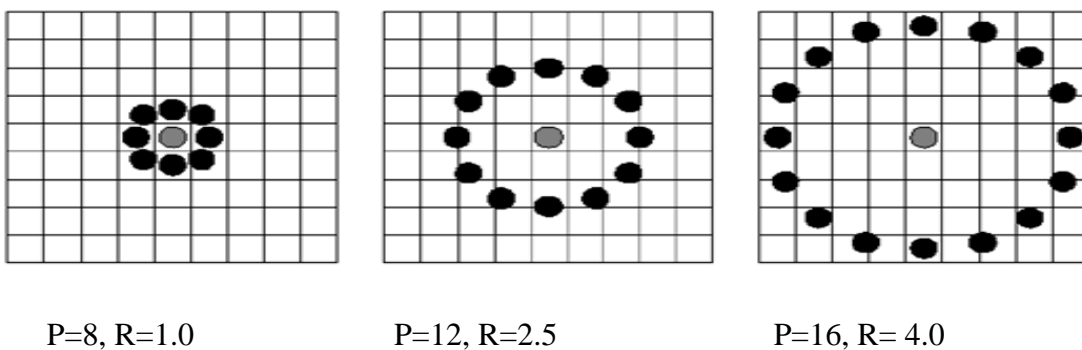


Figure 2.10 : Représentation du pixel central g_c et de son voisinage g_p (Paulhac, 2009).

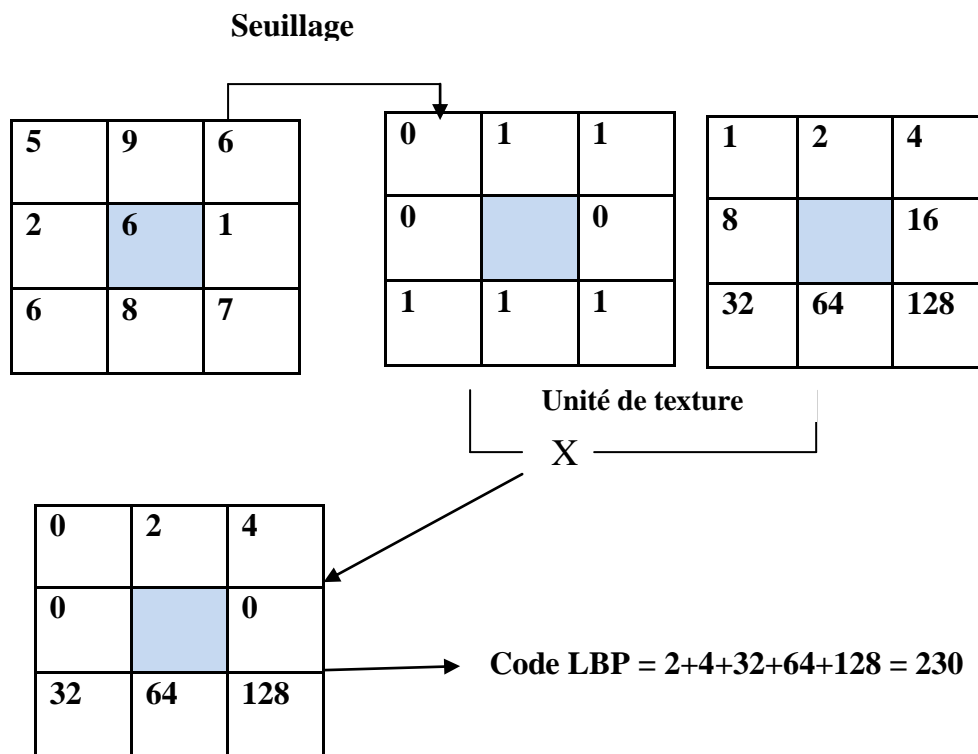


Figure 2.11 : Construction d'un motif binaire et calcul du code LBP (Paulhac, 2009).

Le code LBP est produit en multipliant les valeurs seuils par les poids donnés aux pixels correspondants, et en additionnant le résultat (Mäenpää and Pietikäinen, 2004). La construction du motif binaire et le calcul du code LBP sont présentés dans la figure 2.11.

Comme les méthodes d'analyse de la texture ont traditionnellement été divisées en deux catégories : La première, appelée approche statistique ou stochastique, qui traite les textures comme des phénomènes statistiques, où la formation d'une texture est décrite avec les propriétés statistiques des intensités et des positions des pixels ; La seconde catégorie, appelée approche structurelle, qui introduit le concept de primitives de texture, souvent appelées texels ou textons. La méthode LBP peut être considérée comme une approche véritablement fédératrice. Au lieu d'essayer d'expliquer la formation de la texture au niveau du pixel, des motifs locaux se forment. Chaque pixel est étiqueté avec le code de la primitive de texture qui correspond le mieux au quartier local. Ainsi, chaque code LBP peut être considéré comme un micro-texte local ; les primitives détectées par la LBP comprennent les taches, les zones plates, les bords, les extrémités des bords, les courbes et ainsi de suite. Différentes primitives de texture détectées par le LBP (un pic, un creux, une fin de ligne, une bordure, un coin) sont montrées dans la figure 2.12 où les uns sont représentés par des cercles blancs, et les zéros par des noirs séparément (Mäenpää and Pietikäinen, 2004).

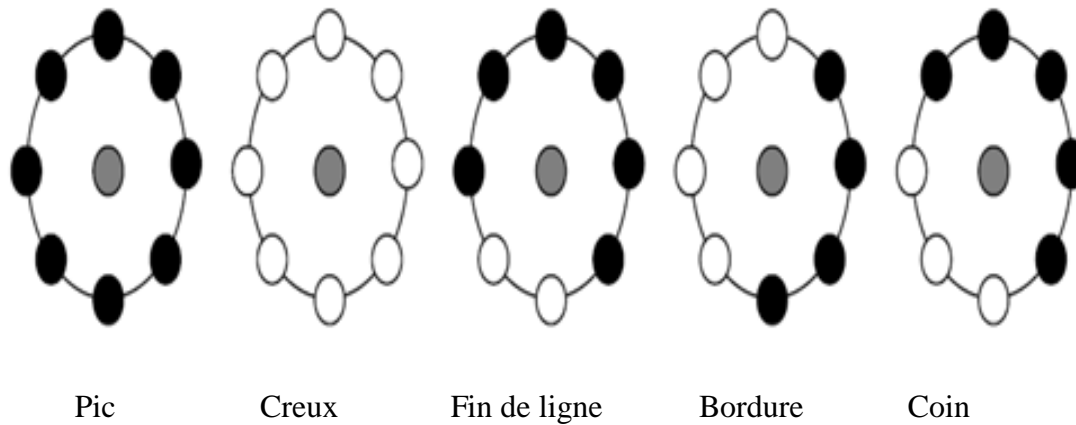


Figure 2.12 : Différentes primitives de texture détectées par le LBP (Mäenpää and Pietikäinen, 2004).

La distribution des LBP présente donc les deux propriétés d'une méthode d'analyse structurelle, les primitives de texture et les règles de placement, d'autre part, la distribution n'est qu'une statistique d'une image filtrée non linéairement, ce qui fait clairement que la méthode est statistique. Pour ces raisons, il faut supposer que la distribution des LBP peut être utilisé avec succès pour reconnaître une grande variété de types de texture, auxquels les méthodes statistiques et structurelles ont été traditionnellement appliquées séparément (Mäenpää and Pietikäinen, 2004).

2.3.2 La classification

La classification est l'une des tâches de prise de décision les plus fréquemment rencontrées dans l'activité humaine. Un problème de classification survient lorsqu'un objet doit être affecté à un groupe ou à une classe prédéfinie en fonction d'un certain nombre d'attributs observés liés à cet objet. De nombreux problèmes industriels sont identifiés comme des problèmes de classification, par exemple, les prévisions boursières, les prévisions météorologiques, les prévisions de faillites, les diagnostics médicaux, la reconnaissance vocale, la reconnaissance de caractères, etc. (Sathya and Annamma, 2013).

Dans les systèmes d'aide au diagnostic, la classification est considérée comme la dernière phase qui différencie et étiquette l'anomalie. Les méthodes de classification jouent un rôle important dans le diagnostic et l'éducation dans le domaine de la médecine. Les approches de classification sont classées en deux groupes (supervisées / non supervisées) (Jalalian et al., 2017).

Les problèmes de classification peuvent être résolus à la fois mathématiquement et de manière non linéaire. La difficulté de résoudre ces problèmes mathématiquement réside dans la précision et la distribution des propriétés des données et des capacités du modèle (Sathya and Annamma, 2013).

La classification supervisée examine un grand nombre de données inconnues et les répartit dans des classes connexes en fonction de leurs caractéristiques ; La principale différence entre les méthodes non supervisées et supervisées est que les méthodes non supervisées n'exigent pas de classe prédéterminée ; Dans une classification supervisée avec succès, toutes les classes doivent être définies et les propriétés de ces classes doivent être extraites pendant la phase de formation ; En général, les techniques de classification supervisées sont utilisées afin de classer les images cliniques (Jalalian et al., 2017). La figure 2.13 résume les techniques de classifications.

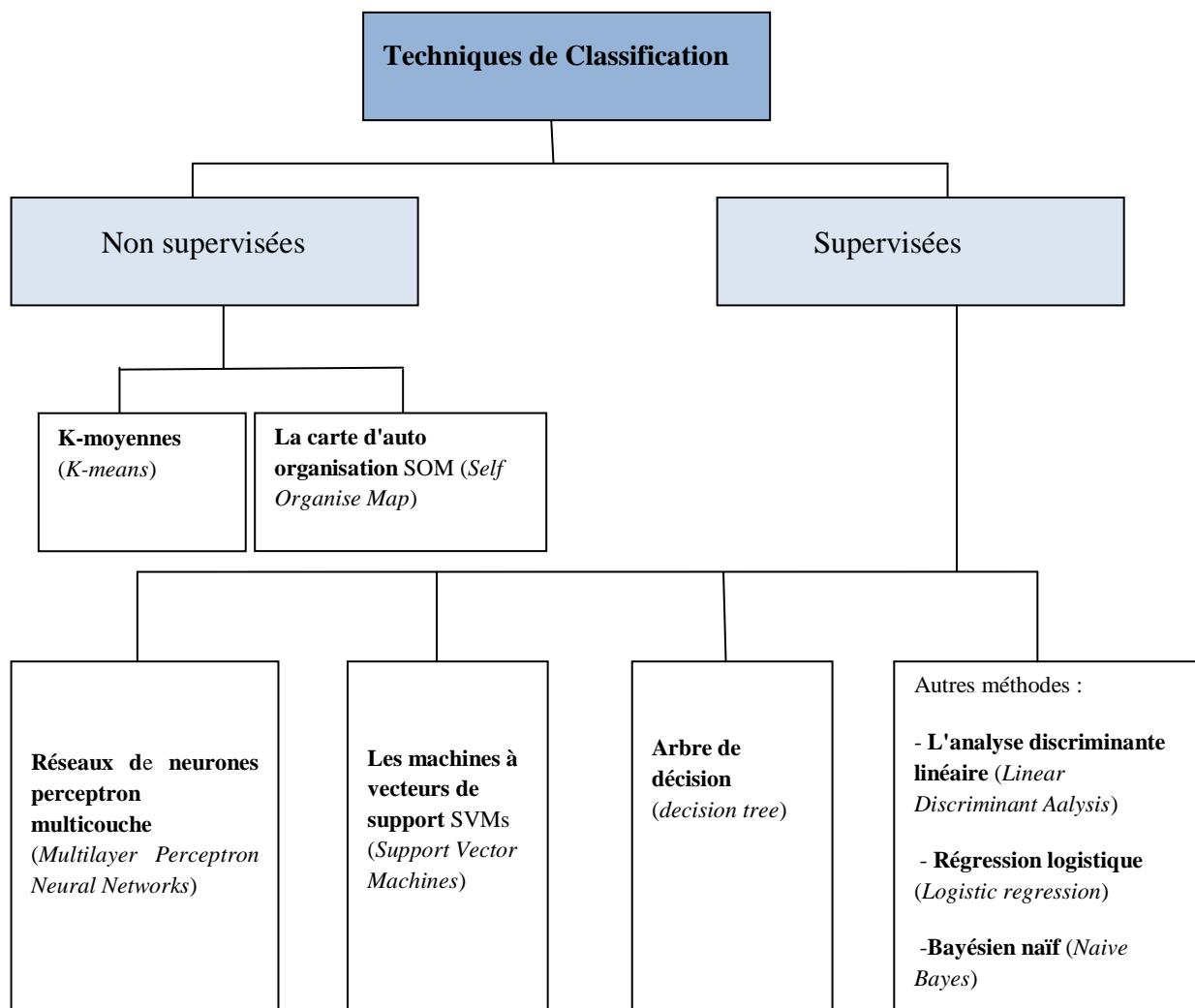


Figure 2.13 : Les techniques de classification (Jalalian et al., 2017).

2.3.2.1 Les méthodes de classifications non supervisées

Dans les méthodes d'apprentissage non supervisées, on considère un ensemble d'échantillons de formation dont seule la valeur d'entrée est spécifiée pour eux et on ne dispose pas d'informations précises sur la sortie (Hemmatian and Sohrabi, 2019).

En général, l'apprentissage non supervisé tente de fonder les décisions de regroupement directement sur les propriétés de l'ensemble d'échantillons donné, sans utiliser des échantillons de formation (Olaode et al., 2014).

Olaode et al. (Olaode et al., 2014) ont résumé les principaux algorithmes d'apprentissage automatique non supervisé nécessaires à la catégorisation non supervisée des images (voir figure 2.14).

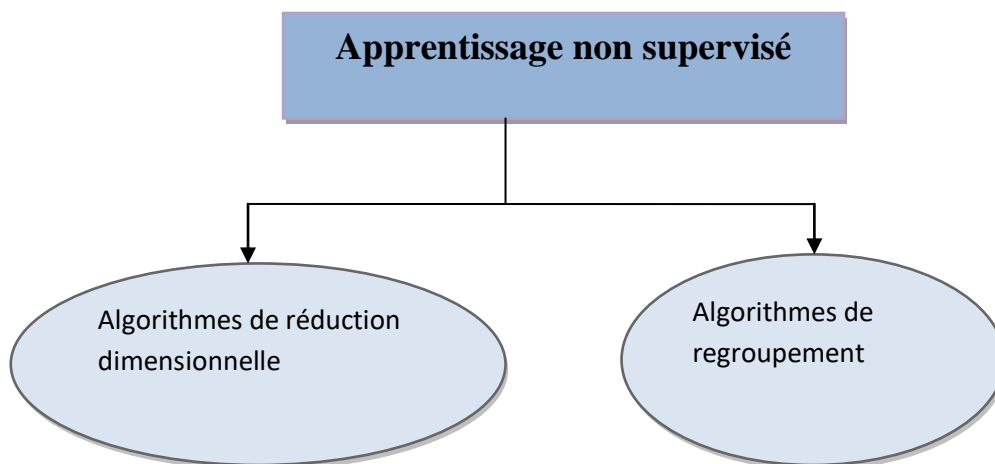


Figure 2.14 : Les approches de classifications non-supervisées

➤ Algorithmes de réduction des dimensions

Il est souvent nécessaire de réduire la dimension des échantillons dans un ensemble de données avant de pouvoir reconnaître les modèles. Par exemple, dans le sac de mots BOV (*Bag-of-Visual words*), qui est une représentation d'image populaire à des fins de classification, qui utilise un histogramme de mots visuels afin de représenter efficacement une image pour les annotations et les tâches de récupération d'images, l'application de la modélisation du BOV peut produire des représentations d'images à 1000 dimensions qui peuvent rendre la catégorisation d'un ensemble d'image représentation BOV inefficaces, surtout lorsqu'il s'agit de traiter une grande collection (1000 échantillons et plus) (Olaode et al., 2014).

Ce problème peut être minimisé par l'estimation d'un modèle global assez grossier, qui caractérise les échantillons de la collection à l'aide des statistiques descriptives, qui réduisent les dimensions des données à l'aide d'un ensemble de variables basé sur les propriétés observées dans les régions de haute densité de probabilité sur l'espace de l'échantillon. Les méthodes d'apprentissage non supervisées les plus courantes permettant d'obtenir des statistiques descriptives comprennent l'analyse en composantes principales (ACP), la factorisation en matrices non-négatives (NMF) et l'analyse en composantes indépendantes (ACI) (Hastie et al., 2008 ; Olaode et al., 2014). Il s'agit des approches linéaires qui ne sont pas souhaitables pour réduire la dimension des images, car les données d'images possèdent des structures compliquées, qui peuvent ne pas être représentées de façon pratique dans un sous-espace linéaire (Li et al., 2009 ; Olaode et al., 2014), par conséquent les cartographier dans un espace linéaire à faibles dimensions peut entraîner une perte significative de la précision de la catégorisation (Olaode et al., 2014).

➤ Algorithmes de regroupement

Les algorithmes de regroupement regroupent les échantillons d'un ensemble de telle sorte que deux échantillons dans le même cluster, sont plus semblables l'un à l'autre par rapport deux autres échantillons du cluster différent. Les méthodes de regroupement peuvent être classées en deux grandes catégories : les méthodes non paramétriques et les méthodes paramétriques. Les méthodes non paramétriques consistent à trouver des groupements naturels (clusters) dans un ensemble de données en utilisant une évaluation du degré de différence (comme la distance euclidienne) entre les échantillons de l'ensemble de données. Pour cela, il faut définir une mesure de (dis) similitude entre les échantillons, définir une fonction de critère pour la mise en grappes et définir un algorithme pour minimiser (ou maximiser) la fonction de critère ; Les algorithmes de regroupement non paramétriques les plus courants sont : les k-means, les algorithmes de regroupement hiérarchiques et les algorithmes de regroupement spectral (Olaode et al., 2014).

Alors que les méthodes paramétriques déduisent la structure du modèle sous-jacent à partir de l'ensemble de données, les approches paramétriques imposent une structure à l'ensemble de données. L'apprentissage paramétrique suppose que les échantillons de l'ensemble de données peuvent être représentés par une fonction de probabilité composée de plusieurs composantes (Hastie et al., 2008 ; Olaode et al., 2014).

Dans les méthodes de regroupement paramétrique, chaque échantillon d'un ensemble est décrit comme une combinaison d'un nombre fini de fonctions et d'échantillons ayant des combinaisons similaires et supposés être dans le même cluster. L'utilisation de la méthode de regroupement paramétrique probabiliste telle que le modèle de mélange gaussien (MGM) (Datta et al., 2008) et le modèle thématique (Mallapragada et al., 2010) s'est avérée efficace dans une grande variété d'applications concernant l'analyse de données continues et discrètes, respectivement (Zhang et al., 2012 ; Olaode et al., 2014).

2.3.2.2 Les méthodes de classification supervisées

Dans l'apprentissage supervisé, le processus d'apprentissage est réalisé en utilisant les données de la formation dans lequel, la valeur de sortie est spécifiée pour toute entrée et le système tente d'apprendre la fonction, en faisant correspondre l'entrée à la sortie, c'est-à-dire en devinant la relation entre l'entrée et la sortie (Hemmatian and Sohrabi, 2019).

Jalalian et al (Jalalian et al., 2017) ont décrit quelques techniques de classification supervisées avec leurs avantages et inconvénients (voir tableau 2.1). D'autres techniques de classification se trouvent en **annexe B**.

Méthode	Description	Avantages	Inconvénients
Réseaux de neurones perceptron multicouche MLPNNs (Multilayer Perceptron Neural Networks)	<ul style="list-style-type: none"> – Le MLP est un modèle de réseau neuronal artificiel à action anticipée qui est utilisé pour concevoir des systèmes de reconnaissance afin de distinguer des modèles spécifiques. – Un ANN (<i>Artificial Neural Networks</i>) est une séquence de plusieurs couches, chaque couche est composée d'un ensemble de neurones. – Tous les neurones d'une couche sont reliés par certains poids aux 	<ul style="list-style-type: none"> – Il s'agit d'un apprentissage non-paramétrique et applicable sur des entrées bruyantes. – Capable de modéliser des modèles complexes non linéaires et des problèmes à haute dimension. – Différentes fonctions du noyau peuvent être 	<ul style="list-style-type: none"> – Dépendance du nombre de couches cachées et du nombre de neurones dans chaque couche. – Le choix du type de réseau est très efficace sur le résultat de la classification.

	<p>neurones de la couche précédente et de la couche suivante.</p> <ul style="list-style-type: none"> – L'efficacité et la précision dépendent du nombre d'entrées et du nombre de couches. 	<p>sélectionnées.</p>	
<p>Les machines à vecteurs de support, SVMs <i>(Support Vector Machines)</i></p>	<ul style="list-style-type: none"> – Un classifieur discriminant pour identifier les hyperplans pour les classes binaires ou les dimensions supérieures. – Les résultats optimaux sont obtenus en trouvant la distance maximale au plus proche point de donnée de chaque classe. – Une faible généralisation d'erreur de la classification peut être obtenue par une marge plus large. 	<ul style="list-style-type: none"> – Il s'agit d'une technique d'apprentissage non paramétrique. – Efficace dans le classifieur à hautes dimensions. – Différentes fonctions du noyau peuvent être spécifiées. 	<ul style="list-style-type: none"> – Les performances et la précision dépendent fortement de la sélection des paramètres du noyau. – Les mauvaises performances avec un plus grand nombre de caractéristiques sont bien plus que le nombre d'échantillons.
<p>Arbre de décision <i>DT (Decision Tree)</i></p>	<ul style="list-style-type: none"> – DT est un outil d'aide à la décision hiérarchique permettant de cloisonner les données à des sous-ensembles uniformes. – Il s'agit d'une fonction 	<ul style="list-style-type: none"> – Il s'agit d'une méthode d'apprentissage non paramétrique utilisée pour la classification et régression. – Simple à 	<ul style="list-style-type: none"> – Extrêmement instable et dépendant des données. – Une petite

	<p>récursive exécutée pour sélectionner le meilleur attribut afin de diviser les données et de les étendre aux nœuds des feuilles.</p>	<p>comprendre et à interpréter.</p> <ul style="list-style-type: none"> – Capable de gérer des problèmes de sorties multiples. 	<p>modification des données entraîne la génération d'un arbre complètement différent.</p>
<p>Analyse discriminante linéaire ou LDA (<i>Linear Discriminant Analysis</i>)</p>	<ul style="list-style-type: none"> – LDA est un algorithme de réduction de la dimensionnalité pour la classification. 	<ul style="list-style-type: none"> – Il peut traiter la non-linéarité de la distribution des échantillons. – Apprend correctement les représentations à faible dimension pour la classification. 	<ul style="list-style-type: none"> – Incapable de trouver la structure non linéaire cachée dans les données à haute dimension. – Problème de singularité matricielle inhérent.

Tableau 2.1 : Description des techniques de classification supervisées (Jalalian et al., 2017).

2.3.2.3 Évaluation des algorithmes de classification

Pour évaluer la performance d'un modèle, l'évaluation doit être effectuée sur des données non vues lors de la formation du modèle. L'un des problèmes liés à l'utilisation d'un ensemble de données complètement différent pour tester et entraîner le modèle est que les informations de l'ensemble de test peuvent contenir des informations importantes qui sont perdues, car elles ne sont jamais utilisées pour entraîner le modèle (Guerra et al., 2011).

Une technique courante qui permet d'évaluer les performances sans perdre d'informations est la validation croisée **k-fold** (Stone, 1974 ; Guerra et al., 2011). L'ensemble de données de

taille m est partitionné aléatoirement en k ensembles ("folds") tous de taille m/k . Ensuite, $k-1$ ensemble sont utilisés pour entraîner le modèle, qui est ensuite évalué sur le seul ensemble non utilisé ; Ce processus est répété k fois, en laissant à chaque fois de côté un ensemble différent pour l'évaluation du modèle. La mesure finale de la performance du modèle est la moyenne des k -exécutions (Guerra et al., 2011).

2.4 La super-résolution

Ces dernières années, avec l'impact du boom d'internet et le développement rapide de l'information, les besoins des gens en matière de traitement du signal et de l'information ont progressivement augmenté, et le traitement de l'image est une partie importante du traitement de l'information (Hou et al., 2020).

La technique de la super-résolution SR (*super-resolution*) des images est particulièrement importante dans le traitement des images, son principe est de passer une ou plusieurs images à basse résolution LR (*low-resolution*) à une image finale à haute résolution HR (*high-resolution*) (Hou et al., 2020).

La méthode de la super-résolution traditionnelle consiste à agrandir la petite image LR à la taille requise par l'interpolation, et utiliser ensuite l'algorithme de reconstruction afin d'obtenir une image HR ; La super-résolution reconstruit l'image qui a des détails clairs et contient beaucoup d'informations, de telle sorte qu'elle peut être largement utilisée (Thornton et al., 2006 ; Shi et al., 2013) dans (Hou et al., 2020). La figure 2.15 illustre l'application de la technique de la super-résolution.

L'obtention d'une image à haute résolution est importante pour plusieurs applications d'imagerie, elle permet d'obtenir des détails possibles sur des objets observés, ce qui permet une meilleure vision de la réalité, par exemple, dans la vidéo de la scène d'une crime, les images HR fournissent des preuves plus favorables aux organes de sécurité publique pour les enquêtes ; dans les applications d'imagerie satellitaire telles que la télédétection et la détection des ressources, les images HR peuvent facilement distinguer deux ou plusieurs objets similaires ; et en médecine, les images HR améliorent considérablement la précision du diagnostic avec une base de diagnostic plus précise. Par conséquent, un algorithme de super-résolution d'image a une grande valeur pour la recherche, et il existe actuellement de nombreux algorithmes de super-résolution (Hou et al., 2020).

Dans les algorithmes de super-résolution qui sont basés sur l'interpolation, l'image LR d'entrée est principalement débruitée, floue et sur-échantillonnée afin d'améliorer la résolution de l'image ; les algorithmes d'interpolation les plus classiques sont : l'interpolation du plus proche

voisin, l'interpolation linéaire, l'interpolation bicubique, et l'interpolation des splines (Temizel and Vlachos, 2006 ; Hou et al., 2020).

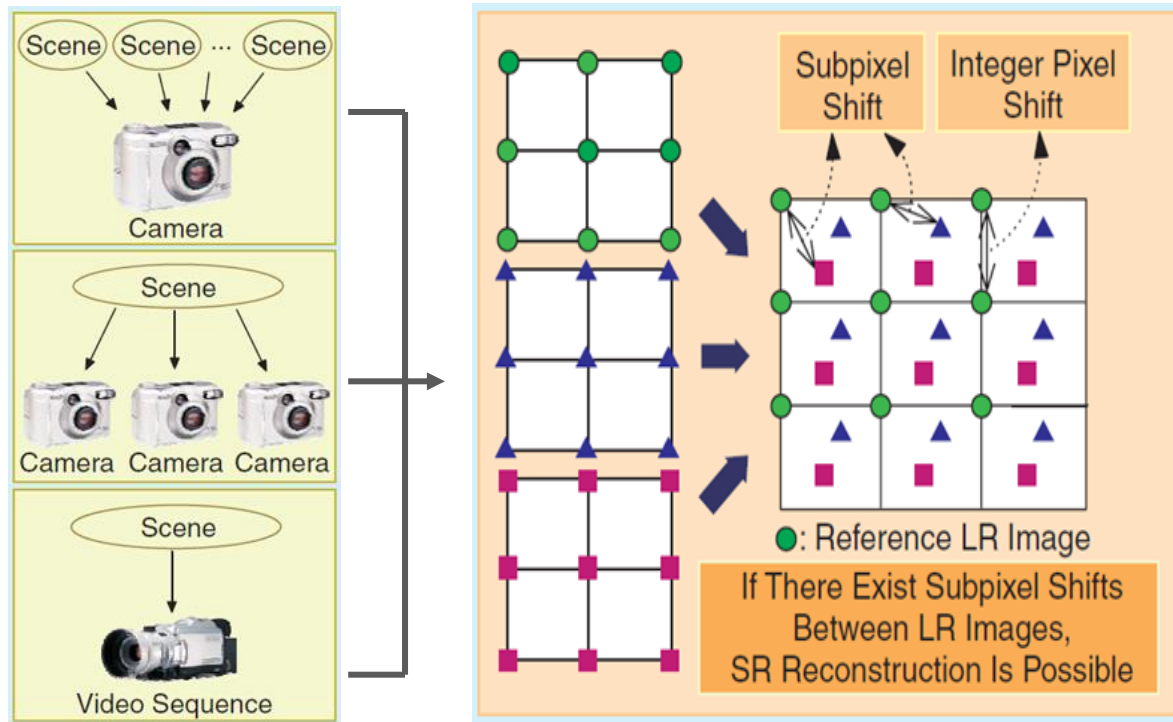


Figure 2.15 : Exemple d'application de la super-résolution (Park et al., 2003).

Le succès de certaines techniques de super-résolution est fondé sur la modélisation du lien entre les images à basse résolution et l'image à haute résolution. Cette modélisation explique comment obtenir des images à basse résolution à partir d'une image à haute résolution, sachant qu'une image à basse résolution est une version dégradée d'une image à haute résolution. Les composantes du processus de dégradation sont : le flou, le bruit, le mouvement et le sous-échantillonnage (Yahiaoui, 2019).

La figure (2.16) modélise le processus d'acquisition d'images (Yahiaoui, 2019), où :

- gk : Ensemble de k images à basse résolution.
- D : Opérateur du sous-échantillonnage.
- Mk : Opérateur de décalage (translation, rotation).
- Bk : Matrice du flou.
- f : Image à haute résolution.
- εk : Bruit du système.

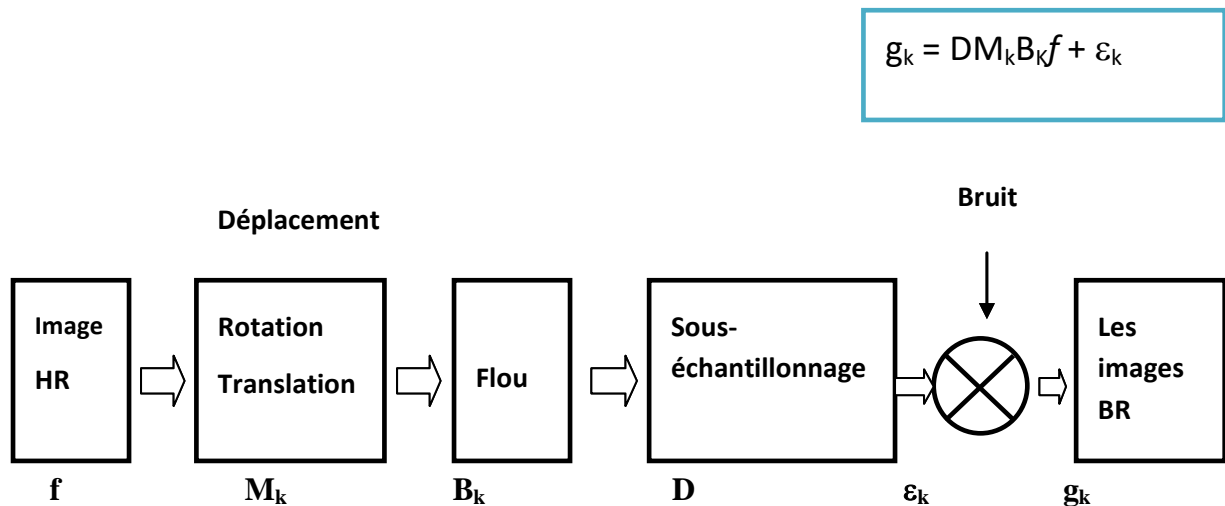


Figure 2.16 : Modélisation du processus d'acquisition d'images (Yahiaoui, 2019).

Les méthodes de la super-résolution qui sont appliquées sur les images peuvent être basées sur des images multiples (voir figure 2.17) ou bien sur une seule image (Zheng et al., 2010).

En super-résolution multi-images, les images à basse résolution dépeignent généralement la même scène, cela signifie que ces images sont également bien déformées que décalées avec une précision subpixel (Park et al., 2003 ; Zheng et al., 2010).

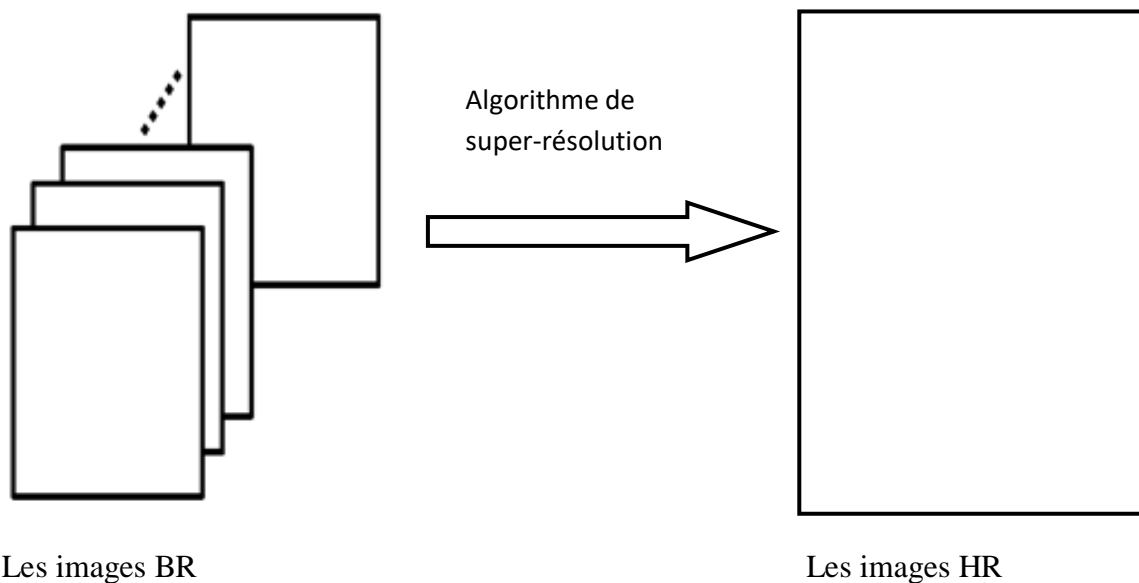


Figure 2.17 : La super-résolution multi-images (Yahiaoui, 2019).

Les techniques de la super résolution des images peuvent être classées principalement en techniques basées sur la reconstruction et en techniques basées sur l'apprentissage (Makwana

et Mehta, 2013) (voir figure 2.18) ; Dans l'approche basée sur l'apprentissage, la relation entre une image à basse résolution et son image à haute résolution correspondante est examinée par le biais d'une paire de patches LR et HR. Les données d'apprentissage sont utilisées afin de prédire l'image à plus haute résolution. L'approche basée sur la reconstruction peut être utilisée dans le domaine fréquentiel ou dans le domaine spatial. La simplicité en théorie est un avantage majeur de l'approche dans le domaine fréquentiel, en outre, l'approche fréquentielle est également pratique pour une mise en œuvre parallèle, toutefois, cette approche offre un peu de souplesse pour ajouter des contraintes a priori, des modèles de bruit et des modèles de dégradation variant dans l'espace. D'autre part, les techniques du domaine spatial sont plus souples dans l'incorporation de contraintes a priori et ont de meilleures performances dans les images reconstruites, néanmoins, ces méthodes présentent également des inconvénients tels qu'un travail théorique compliqué et une charge de calcul relativement importante (Makwana et Mehta, 2013) ; Une comparaison des deux principales classes de techniques de super-résolution dans le domaine fréquentiel et le domaine spatial, se trouve dans le tableau 2.2.

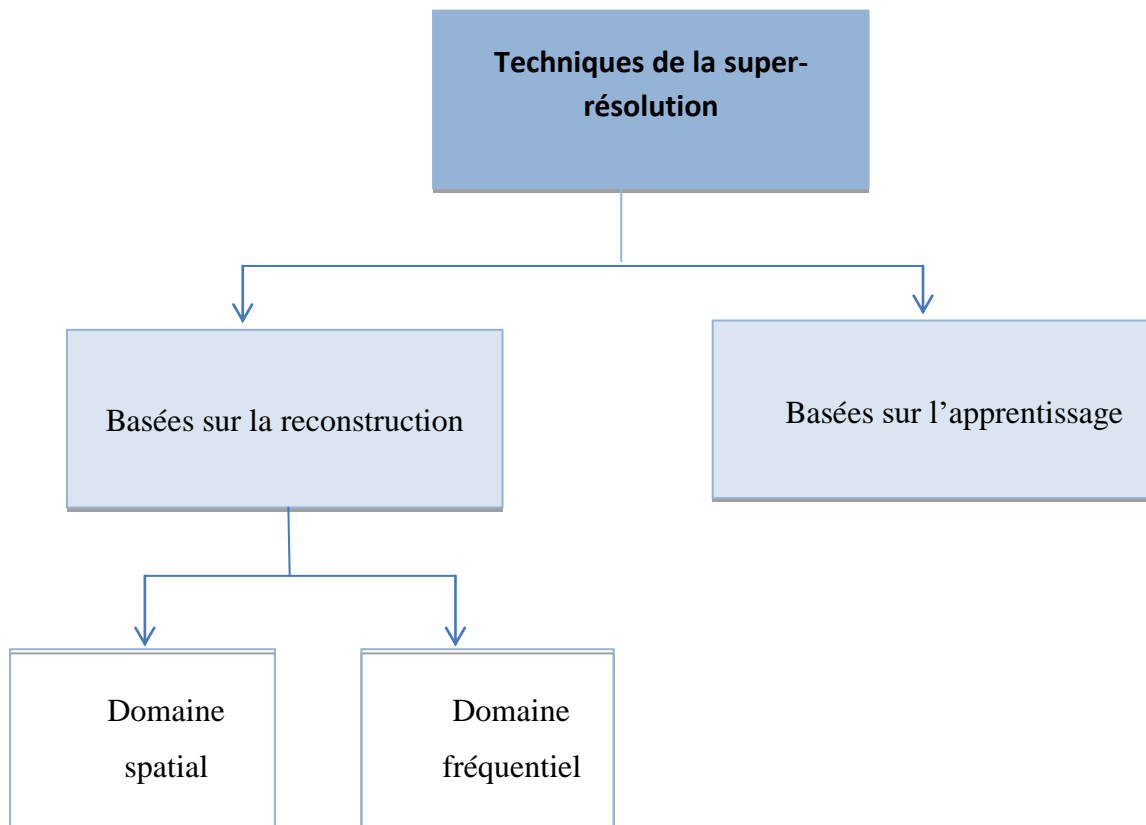


Figure 2.18 : Classification des techniques de la super-résolution

	Domaine fréquentiel	Domaine spatial
Domaine d'observation	Domaine de fréquence	Domaine spatial
Modèle de mouvement	Translation globale	Presque illimité
Modèle de dégradation	Translation globale	Presque illimité
Modèle de bruit	Limité	Presque illimité
Informations a priori	Limitées	Presque illimitées
Simplicité	Très simple	Généralement complexe
Coût de calcul	Faible	Haut
Régularisation	Limitée	Excellente
Extensibilité	Pauvre	Excellente
Applicabilité	Limitée	Large
Performance	Bonne pour une application spécifique	Bonne

Tableau 2.2 : Comparaison entre les techniques de la super résolution du domaine fréquentiel et du domaine spatial (Makwana et Mehta, 2013).

2.5 Conclusion

Ce chapitre a permis de brosser un portrait global des différentes notions de base sur l'extraction des caractéristiques, la classification et la super-résolution des images, afin de faire des choix appropriés au futur pour améliorer la qualité des images mammographiques, et pour caractériser et classer les régions d'intérêt dans ces images.

Dans le prochain chapitre, nous exposerons les travaux de classification des régions d'intérêt dans les images mammographiques, ainsi que l'application de la technique de la super-résolution sur les images médicale.

Chapitre 03

Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales

3.1 Introduction

Le présent chapitre est divisé en trois parties. Au niveau de la première partie, nous donnons un rappel sur la classification automatique des lésions mammaires.

Dans la deuxième partie, nous dressons un état de l'art sur les travaux récents d'extraction de caractéristiques et de classification automatique utilisés dans les systèmes d'aide au diagnostic, et qui sont appliqués sur les régions suspectes dans les images mammographiques. Nous soulignons l'importance de l'utilisation de la méthode de la super-résolution dans le domaine médicale, au niveau de la dernière partie.

3.2 Les systèmes d'aide au diagnostic pour la classification des lésions mammaires

La détection précoce du cancer du sein est très importante, car elle peut réduire considérablement les taux de mortalité (Oeffinger et al., 2015). L'examen de dépistage donne la possibilité de détecter le cancer du sein à un stade précoce, par conséquent, le traitement aura plus de chances pour réussir (Valarmathie et al., 2016). Il existe plusieurs techniques d'imagerie médicale pour l'examen des seins, telles que l'imagerie par résonance magnétique, l'imagerie par ultrasons, l'imagerie par rayons X, etc. La mammographie est une procédure radiologique spéciale pour la détection du cancer du sein qui est la meilleure et la moins chère de toutes les autres techniques, avec moins d'effets secondaires dans la détection des tumeurs qui provoquent le cancer du sein (Valarmathie et al., 2016). La lecture des images mammographiques est un travail très exigeant pour les radiologues (Yadollahpour et Shoghi, 2015). La figure 3.1 montre cette difficulté sur deux images mammographiques extraites de la base de données MIAS (*Mammographic Image Analysis Society*).

C'est pourquoi, l'utilisation des systèmes de diagnostic assisté par ordinateur est très utile, afin de donner un deuxième avis aux radiologues.

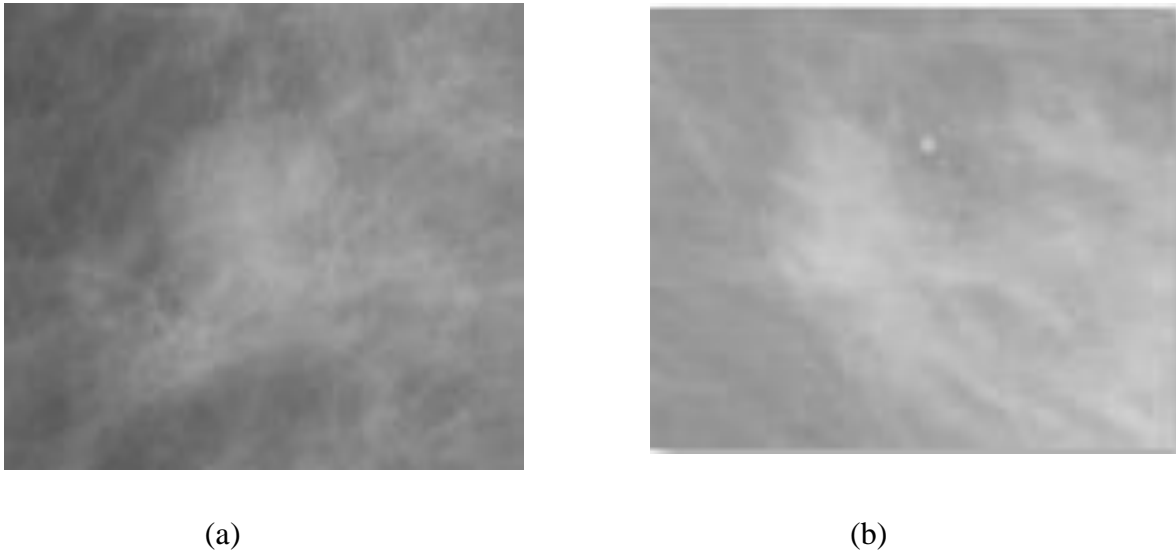


Figure 3.1 : Deux types de masses (a) une masse bénigne et (b) une masse maligne (web 08).

Les phases essentielles des systèmes de diagnostic assisté par ordinateur dédiés aux images mammographiques sont (Cheikhrouhou, 2012) (voir figure 3.2) :

- Phase de prétraitement : a pour but l'amélioration de la qualité de l'image avant toutes utilisations.
- Phase de segmentation : permet de cerner la lésion d'intérêt.
- Phase d'extraction de descripteurs : a pour but de caractériser les lésions à travers des formulations mathématiques.
- Phase de classification et de prise de décision en utilisant un classifieur donnée (Cheikhrouhou, 2012).

Depuis plusieurs années, beaucoup de travaux ont été faits afin de concevoir des systèmes d'aide au diagnostic capable de cerner des régions d'intérêt dans les images mammographiques et de faire la classification des parties trouvées. Eddaoudi (Eddaoudi, 2012) a cité quelques systèmes SAD commercialisés comme :

- «Image Checker M1000», de la société R2 technology développé à l'Université de Chicago.
- Second Look, de la société CADX Medical Systems.
- «Mammex TR», de la société Scanis.
- «GP-CALMA» (Grid Platform for Computer Assited Library for Mammography), développé principalement en Italie et dont l'originalité focalise sur la conception d'une grande base de données d'images (plus de 5000 images) (Eddaoudi, 2012).

Le principe de ces systèmes est généralement le même, et le développement de tels systèmes n'ignore jamais le rôle d'un médecin, car c'est un outil qui l'aide dans l'interprétation.

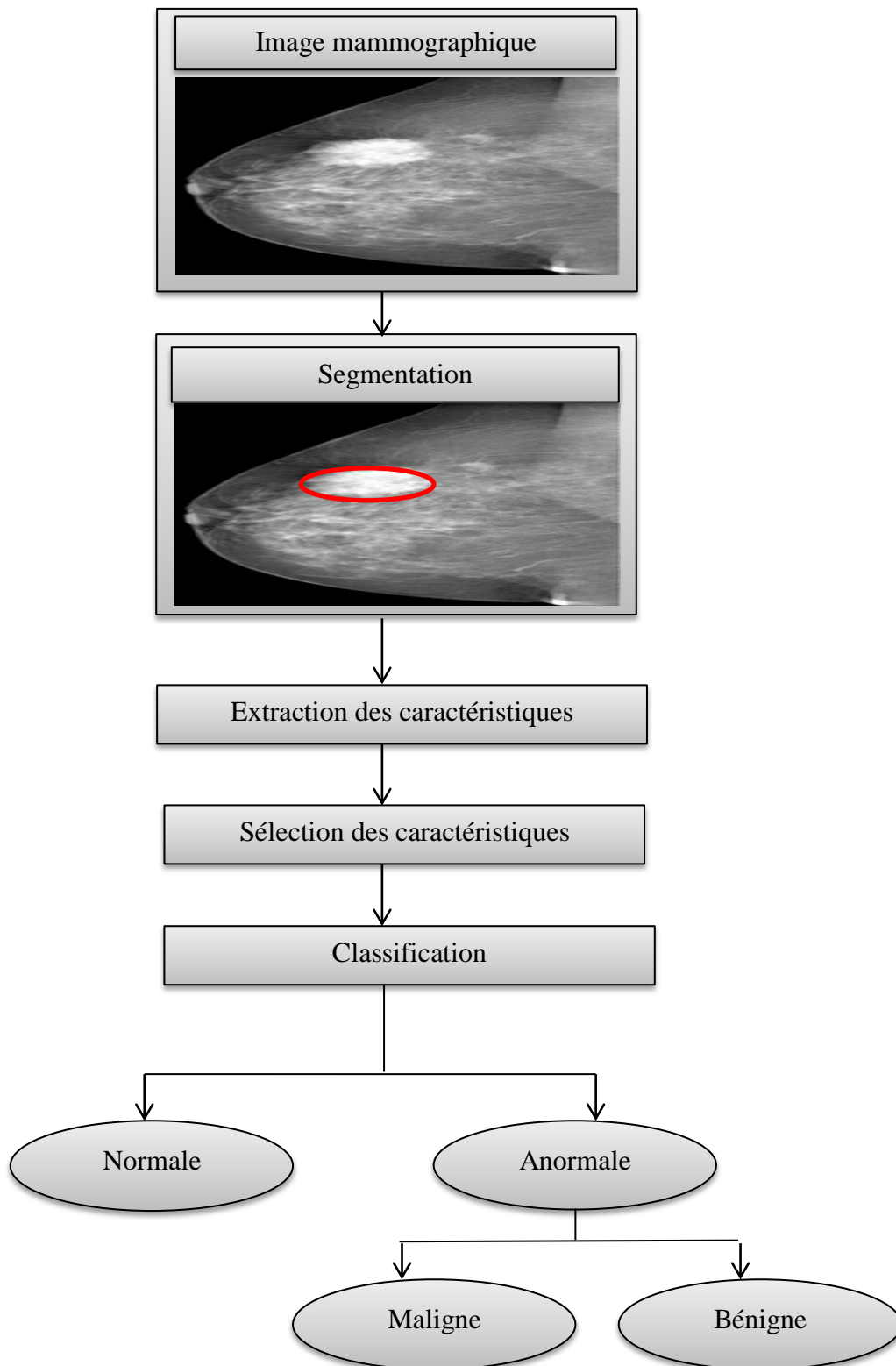


Figure 3.2 : Schéma général d'un système d'aide au diagnostic dédié aux images mammographiques

3.3 Synthèse des travaux de classification automatique des lésions mammaires

Ces dernières années, la reconnaissance des formes, qui implique l'extraction de caractéristiques et la classification des masses, a attiré l'attention de nombreux chercheurs.

En général, la reconnaissance des formes peut être réalisée par des algorithmes classiques basés sur des approches de sélection de caractéristiques ou par des algorithmes récents basés sur des approches d'apprentissage profond (voir figure 3.3).

Dans les approches conventionnelles, les modèles sont généralement basés sur des caractéristiques prédéfinies, qui nécessitent une détermination préventive des caractéristiques qui fonctionnent bien pour la tâche en question ; Par contre, dans les approches d'apprentissage profond, une prédéfinition des caractéristiques d'image n'est pas nécessaire, car elles sont apprises automatiquement à partir des données étiquetées ; L'apprentissage profond permet de déterminer des caractéristiques nuancées par l'algorithme d'apprentissage pour la tâche ciblée, où les caractéristiques intrinsèques qui peuvent ne pas être identifiables par l'évaluation visuelle humaine peuvent être automatiquement identifiées et utilisées pour l'interprétation des images (Aboutalib et al. 2018).

Dans cette section, nous reprenons certains travaux sur l'extraction de caractéristiques et la classification des lésions dans les images mammographiques ; dans le tableau 3.1, nous résumons les travaux de classification automatique des lésions mammaires.

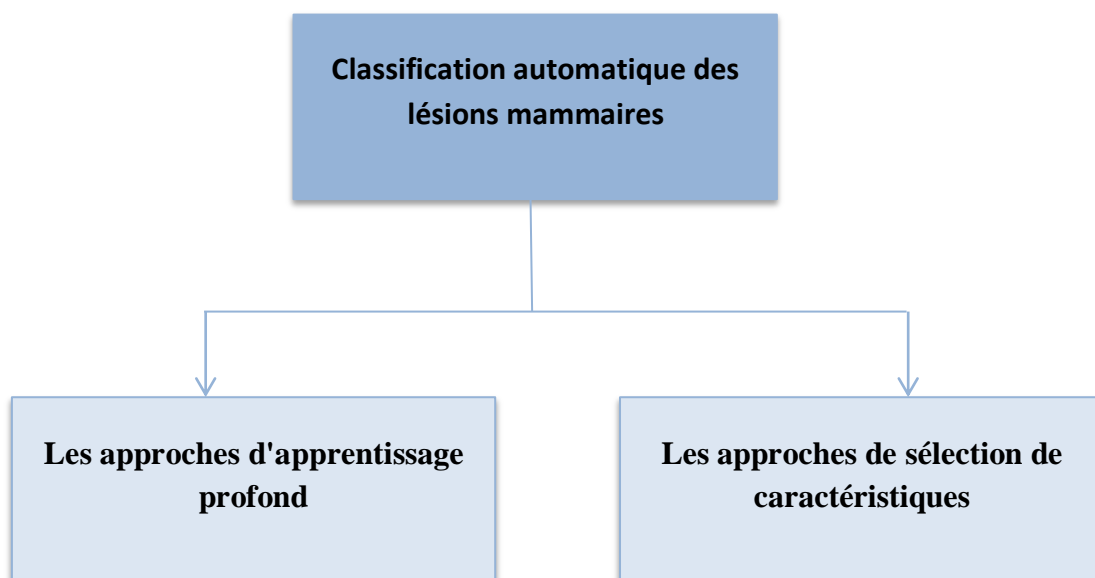


Figure 3.3 : Les approches de classification automatique des lésions mammaires

3.3.1 Classification basée sur les approches de sélection de caractéristiques

La reconnaissance des formes par des approches de sélection de caractéristiques nécessite une description de masse, en utilisant une méthode de sélection de caractéristiques et un algorithme d'extraction de caractéristique ; l'objectif des travaux qui sont basés sur des approches de sélection de caractéristiques, est de trouver la meilleure combinaison de caractéristiques qui donne de bons résultats de classification.

La taxonomie des caractéristiques est basée sur les dimensions des descripteurs de caractéristiques utilisant trois axes : « forme et motif », « spectre » et « densité » (Jalalian et al., 2017) (voir figure 3.4).

De nombreuses recherches ont été menées afin d'améliorer les techniques qui permettent d'extraire les caractéristiques significatives des images mammographiques. Ces recherches sont différentes dans le type de caractéristiques utilisées pour la détection de masse, et dans la stratégie qui a été adoptée pour extraire ces caractéristiques ; Différents types de caractéristiques ont été utilisés ; comme l'intensité, le gradient, la texture, la forme (Xie et al., 2015 ; Khan et al., 2017), les polynômes de Zernike, le filtre de Gabor, l'ensemble des banques de filtres et la transformée en ondelettes discrète, qui pourraient améliorer les performances de la classification (Raghavendraa et al., 2016).

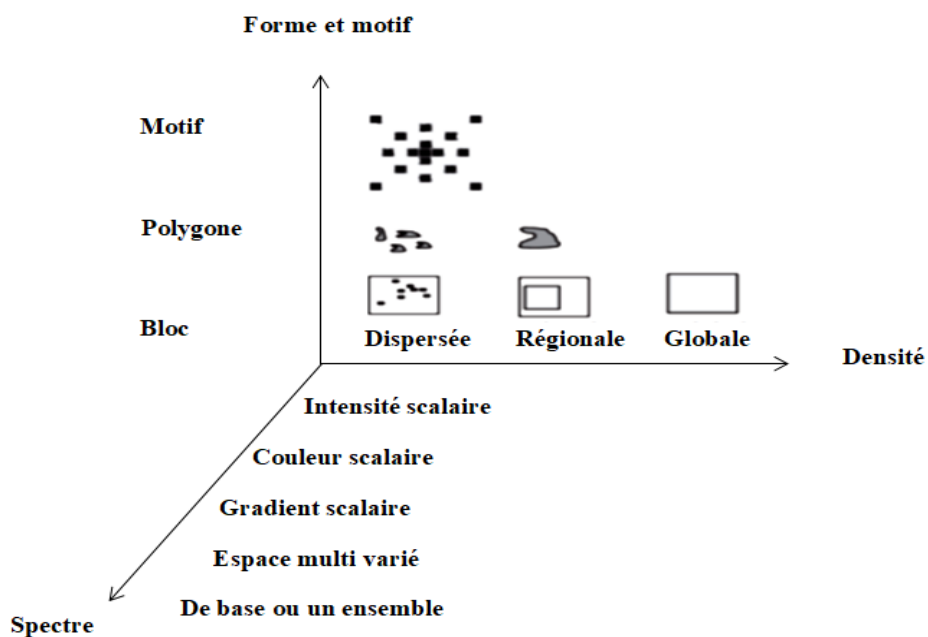


Figure 3.4 : Taxonomie pour les dimensions des descripteurs de caractéristiques (Krig, 2014).

Le processus d'extraction de caractéristiques est suivi généralement d'une étape de classification pour classer les lésions en cas normaux ou anormaux (Raghavendraa et al., 2016).

Les techniques de classification les plus utilisées sont les suivantes : k plus proche voisin, réseau de neurones artificiels, flou, machine à vecteurs de support (SVM) et l'analyse discriminante linéaire (Raghavendraa et al., 2016).

Certains travaux récents qui sont basés sur les approches de sélection de caractéristiques pour classer les lésions mammaires sont présentés dans cette section :

Dans le but de trouver la meilleure combinaison de caractéristiques, Jehlol et al. (Jehlol et al., 2015) ont testé les caractéristiques de texture (les statistiques du premier et du deuxième ordre) avec cinq algorithmes d'apprentissage différents. Ils ont réalisé leur évaluation sur la base de données MIAS. La meilleure précision a été obtenue avec les descripteurs statistiques du second ordre et le classifieur 'forêt aléatoire', avec une précision de 85,8% pour la classification des régions bénigne, maligne et normale.

Görgel et al. (Görgel et al., 2015) ont appliqué la transformée en ondelettes stationnaire sur les régions d'intérêt, ils ont ensuite extrait des caractéristiques basées sur la forme et le niveau de gris avec quelques coefficients d'échelle. Pour l'action de classification, ils ont utilisé une machine à vecteurs de support. Ils ont obtenu une précision de classification de 91,4% et 90,1% en utilisant l'ensemble de données acquises de l'hôpital de l'Université d'Istanbul en Turquie et la base de données MIAS, respectivement.

Khan et al. (Khan et al., 2016) ont appliqué la méthode d'optimisation par essais particuliers afin d'optimiser les banques de filtres de Gabor. Ils ont extrait les caractéristiques de Gabor en utilisant les banques de filtres de Gabor optimisés. Ils ont évalué leur approche avec le classifieur SVM sur 768 régions d'intérêt tirés de la base de données (DDSM). Ils ont obtenu une précision de 93,95% pour la classification des masses.

Singh and Singh (Singh and Singh, 2016) ont proposé un système de classification assisté par ordinateur afin de classer les régions d'intérêt dans les images mammographiques en classes normale, bénigne et maligne, les images sont tirées de la base de données DDSM. Ils ont extrait des descripteurs de texture en utilisant la méthode de la matrice de co-occurrence des niveaux de gris (GLCM) avec différentes valeurs de distance inter-pixel variant de un à huit. Pour l'étape de classification, ils ont opté le classifieur SVM. Le résultat de leur étude a

prouvé que les caractéristiques de GLCM calculées conduisaient à une précision de classification globale maximale de 75%.

Seryasat and Haddadnia (Seryasat and Haddadnia, 2017) ont basé sur les caractéristiques de texture et les caractéristiques morphologiques. Ils ont combiné un classifieur faible avec un classifieur fort dans la phase de classification. Ils ont obtenu une précision de 93% pour l'ensemble de données MIAS et de 90% pour l'ensemble de données DDSM.

Sheba and Raj (Sheba and Raj, 2018) ont utilisé les caractéristiques extraites de la matrices de longueurs de plages 'GLRLM', les caractéristiques de forme, les caractéristiques de GLCM et les caractéristiques statistiques du premier ordre. Ensuite, ils ont sélectionné les caractéristiques optimales en utilisant l'arbre de régression. La classification a été effectuée à l'aide de réseau de neurones artificiels, et ils ont obtenu une précision de 96%.

Braz et al. (Braz et al., 2018) ont diversifié l'analyse des régions de masse par les descripteurs de la géométrie géostatistique. Ils ont évalué chaque extraction de caractéristiques à l'aide de la machine à vecteurs de support. Ils ont obtenu 97,30% du taux de détection pour la base de données MIAS et 91,63% pour la base de données DDSM.

Pezeshki et al. (Pezeshki et al., 2019) ont obtenu une précision de 91,37% et de 93,22% en utilisant respectivement les bases de données mini-MIAS et DDSM, ils ont basé sur la texture, les caractéristiques d'intensité, les caractéristiques de forme et de marge, et le classifieur SVM pour classer les masses comme malignes ou bénignes.

Certains auteurs ont effectué une stratégie de comparaison afin de sélectionner les caractéristiques les plus appropriées :

En utilisant 250 images mammographiques sélectionnées de la base de données DDSM, Nithya and Santhi (Nithya and Santhi, 2011) ont comparé trois types de caractéristiques de texture : les caractéristiques de l'histogramme d'intensité, les caractéristiques du GLCM, les caractéristiques d'intensité. Ils ont prouvé que l'utilisation des caractéristiques de GLCM avec le classifieur de réseau de neurones donnent la meilleure précision qui égale 98%.

Rabidas et al. (Rabidas et al., 2016) ont comparé les caractéristiques du motif binaire local (LBP), LBP variance et LBP complète ; les caractéristiques sélectionnées sont utilisées pour classer les masses comme bénignes ou malignes dans les images mammographiques de la base de données DDSM. La méthode de Rabidas et al. (Rabidas et al., 2016) a atteint une

précision de 92,25%, lorsqu'ils ont sélectionné l'ensemble optimal de caractéristiques en utilisant la méthode de régression logistique suivie d'une classification par l'analyse discriminante linéaire de Fisher.

Uyun et al. (Uyun et al., 2013) ont calculé les caractéristiques statistiques du premier ordre et les caractéristiques statistiques de second ordre à partir des images mammographiques, tirées de la base de données MIAS. Ils ont obtenu le meilleur résultat lorsqu'ils ont utilisé les caractéristiques GLCM dans toutes les directions (0° , 45° , 90° , 135°) et dans l'étape de classification, ils ont basé sur le classifieur 'perceptron multi-couches'.

Bovis and Singh (Bovis and Singh, 2000) ont adopté une approche similaire. Ils ont utilisé les matrices de cooccurrence pour l'extraction des caractéristiques de texture et, ils ont appliqué le réseau de neurones artificiel (ANN) pour la classification. Après avoir testé leur approche sur la base mini-MIAS, ils ont atteint un taux de reconnaissance moyen de 77%.

Prasad and Basha (Prasad and Basha, 2018) ont concentré sur quatre méthodes d'extraction de caractéristiques, telles que les motifs binaires locaux, l'histogramme de gradient orienté (HOG), Gabor, la matrice de co-occurrence de niveau de gris (GLCM), et les tester sur les classifieurs SVM pour la détection du cancer du sein, et ils ont noté que la méthode de texture GLCM est la plus adaptée avec le classifieur SVM, et même la classification est plus précise que les autres méthodes avec une précision de 92%.

Pratiwia et al. (Pratiwia et al., 2015) ont utilisé des caractéristiques de texture. Toutes les caractéristiques sont calculées à partir de la matrice de cooccurrence des niveaux de gris (GLCM). Les caractéristiques de texture sont également le choix de Shanthi and Murali (Shanthi and Murali, 2014). Mais, leur solution est basée sur les filtres de Gabor et l'analyse fractale pour identifier l'apparition la plus fréquente du cancer du sein, à savoir les micro-calcifications, les masses et les déformations architecturales. Avec 192 images utilisées de la base MIAS, ils ont obtenu une précision de classification de 98.44%.

Le concept fractal est également adopté par Saraswathi and Srinivasan (Saraswathi and Srinivasan, 2016) ; ils ont proposé une approche fractale par courbes. La transformée en curvelet similaire est utilisée dans le travail de Saraswathi et al. (Saraswathi et al., 2015) , et à fin de réduire la complexité du temps et de sélectionner les caractéristiques importantes, un algorithme d'optimisation par essais a été proposé. Ensuite, les caractéristiques

sélectionnées sont données au classifieur SVM, ils ont obtenu un taux de classification de 96%.

Pour Liu et Tang (Liu and Tang, 2014), les caractéristiques de texture ne sont pas suffisantes, ils ont donc ajouté des caractéristiques géométriques dans leur proposition. D'autres types d'éléments sont également utilisés, comme des mesures d'énergie de texture proposées par Dheeba et al. (Dheeba et al., 2014), la décomposition en valeurs singulières et la transformation des traces implémentées par Bhanumathi and Suresh (Bhanumathi and Suresh, 2015).

Rouhi et al. (Rouhi et al., 2015) ont basé sur la forme, l'intensité et les caractéristiques de texture ; pour la classification, ils ont utilisé le perceptron multi-couche (MLP).

Pak et al. (Pak et al., 2015) ont basé sur différentes caractéristiques comme la forme des lésions, la surface, la compacité, l'excentricité du moment central, l'étalement, les descripteurs de niveau de gris ; pour l'étape de classification, ils ont utilisé le classifieur AdaBoost pour déterminer la probabilité des masses bénignes et malignes.

Berbar (Berbar, 2018) a proposé trois méthodes hybrides appelées ondelettes-CT1 (*Contourlet Transform*), ondelettes-CT2 et ST-GLCM, où ST-GLCM fusionne sept caractéristiques statistiques et sept caractéristiques de texture extraites de la matrice de cooccurrence de niveau de gris (GLCM), il a utilisé le classifieur SVM.

Herwanto and Arymurthy (Herwanto and Arymurthy, 2013) ont proposé une méthode de classification de la masse et de la micro-calcification dans la mammographie, en utilisant une technique d'association. Cette méthode comprend trois phases : (1) une phase de prétraitement pour améliorer la qualité de l'image, et elle est suivie d'une segmentation de la ROI, (2) une phase pour l'extraction d'une table transactionnelle, et (3) une phase pour l'organisation des règles d'association résultantes dans un modèle de classification. L'ensemble de données mammographiques MIAS a été utilisé pour évaluer la méthode proposée. Les caractéristiques moyennes et les caractéristiques GLCM sont utilisés. Ils ont obtenu une précision de 83%.

La méthode proposée par Jasmine et al. (Jasmine et al., 2013) vise à utiliser la transformée en contourlette non sous échantillonnée pour classifier les régions d'intérêt étant normales ou anormales (bénigne ou maligne). Ils ont utilisé une machine à vecteur de support. Dans leur étude, les images mammographiques de la base de données MIAS ont été utilisées, et ils ont

obtenu une précision de 98,5% pour la classification normale et anormale et 96.15% pour la classification bénigne et maligne.

Le GLRLM, qui est une caractéristique de texture d'ordre supérieur, a été combinée aux caractéristiques du GLCM dans les travaux de Mohanty et al. (Mohanty et al., 2011) afin de distinguer les masses malignes des masses bénignes, ils ont appliqué la technique d'extraction par règles d'association. Ils ont obtenu une précision de 94.9%.

Sakikala and Ezhilarassi (Sakikala and Ezhilarassi, 2016) ont combiné les caractéristiques du GLRLM avec un ensemble de caractéristiques différents : les statistiques du premier ordre, GLCM, Gabor, fractale et les pyramides orientables, avec le classifieur perceptron multicouches, ils ont obtenu une précision de 92,6% pour la base de données DDSM et de 89,1% pour la base de données In Breast.

Dong et al. (Dong et al., 2015) ont ajouté aux caractéristiques statistiques du premier ordre des caractéristiques de forme et de marge. En utilisant un classifieur forêt aléatoire, ils ont obtenu une précision de 97,73%.

Valarmathie et al. (Valarmathie et al., 2016) ont également utilisé les caractéristiques de forme et de marge dans leur combinaison. Ils ont obtenu une précision de 98% en utilisant le classifieur perceptron multicouche.

3.3.2 Classification basée sur les approches d'apprentissage profond

Les progrès récents en matière d'apprentissage profond ont prouvé que les réseaux de neurones profonds formés de bout en bout peuvent apprendre de puissantes représentations de caractéristiques et surpasser les classifieurs construits à partir des caractéristiques extraites (Vincent et al. 2010 ; Kraus et al. 2016), c'est pourquoi plusieurs études ont examiné l'apprentissage profond dans les tâches liées à l'imagerie mammographique du sein, comme la différenciation entre les masses bénignes et malignes. Cette section présente quelques approches de reconnaissance des formes basées sur le principe de l'apprentissage profond.

Pour démontrer que les réseaux de neurones à convolution profonde (DCNN) ont un potentiel dans le SAD dédié aux images mammographiques, Suzuki et al. (Suzuki et al., 2016) ont construit une architecture DCNN, qui se compose de cinq couches convolutionnelle et de trois couches entièrement connectées. Tout d'abord, le DCNN a été formé sur des images naturelles. Ensuite, après la dernière modification de la couche entièrement connectée, la nouvelle DCNN a été formée en utilisant des images mammographiques pour deux catégories

de classification : masse et normale. Les résultats expérimentaux ont montré un taux de classification d'environ 89,9%.

Duraisamy and Emperumal (Duraisamy and Emperumal, 2017) ont introduit un algorithme d'apprentissage profond qui se base sur les réseaux de neurones à convolution (DL-CNN) afin d'apprendre les caractéristiques des groupes de masse et de micro-calcification. Les résultats expérimentaux ont montré que la méthode proposée permet de classer les régions d'intérêt dans les images mammographiques comme normales, bénignes ou malignes avec une précision de 99% en utilisant l'ensemble de données MIAS.

Ribli et al. (Ribli et al., 2018) ont développé un système automatique de détection et de classification des lésions sur une mammographie basé sur le R-CNN plus rapide. Plus rapide R-CNN est basé sur un réseau neuronal à convolution avec des composants supplémentaires pour détecter, localiser et classer les objets dans une image. Ils ont utilisé 16 couches de R-CNN et la dernière couche peut détecter des lésions bénignes ou malignes, où les images mammographiques sont tirées de la base de données DREAM, ils ont atteint un taux de classification d'environ 85%.

Arevalo et al. (Arevalo et al., 2016) ont proposé un modèle CNN qui apprend les caractéristiques des masses avant de les donner à un étape de classification. Au lieu de concevoir des caractéristiques particulières pour expliquer le contenu d'images mammographiques, cette approche les apprend directement à partir des données de manière contrôlée. Les chercheurs ont augmenté la performance de classification jusqu'à 82.6% en termes de mesure AUC. Ils ont utilisé 736 images de l'ensemble de données BCDR F03.

Jadoon, et al. (Jadoon et al., 2017) ont proposé un modèle qui vise une série de trois classes qui sont normales, malignes et bénignes, en utilisant deux méthodes, à savoir CNN-DW et CNN-CT. Ils ont amélioré la technique de classification pour les grands ensembles de données des images mammographiques en utilisant CNN. Ils ont montré que les caractéristiques discriminantes et descriptives peuvent être performantes avec différentes ondelettes, si elles sont utilisées en fonction de leur modèle proposé en combinaison avec CNN. Ils ont utilisé le classifieur SVM pour l'ensemble de données IRMA et ils ont atteint une meilleure précision qui égale 83,74%.

Chapitre 03 : Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales

Jiao et al. (Jiao et al., 2016) ont obtenu une précision de 96,7% en classant les masses mammaires entre bénignes et malignes qui ont été tirés de la base DDSM. Ils ont utilisé un CNN comme extracteur de caractéristiques et un SVM comme classifieur.

Abdel-Zaher and Eldeib (Abdel-Zaher and Eldeib, 2016) ont développé un classifieur en utilisant un réseau de croyance profond non supervisé, suivi par une rétro propagation de gradient supervisé, ce modèle a été testé sur l'ensemble de données Wisconsin, ils ont obtenu une précision de 99.68%.

Référence	Approche basée sur		Base d'images	Précision/ Sensibilité/AUC
	La sélection de caractéristiques	L'apprentissage profond		
(Bovis and Singh, 2000)	√		Mini-MIAS	77%
(Nithya and Santhi, 2011)	√		DDSM	98%
(Mohanty et al., 2011)	√		DDSM	94.9%
(Uyun et al., 2013)	√		MIAS	-
(Herwanto and Arymurthy, 2013)	√		MIAS	83%
(Shanthi and Murali, 2014)	√		MIAS	98.44%
(Liu and Tang, 2014)	√		DDSM	96.15%
(Dheeba et al., 2014)	√		Clinical-database (216 images)	94.16%
(Bhanumathi and Suresh, 2015)	√		MIAS	94.94%
(Jehlol et al., 2015)	√		MIAS	85.8 %

Chapitre 03 : Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales

(Görgel et al., 2015)	√		MIAS/Images de l'hôpital UI	90,1% / 91.4%
(Pratiwia et al., 2015)	√		MIAS	94.29%
(Saraswathi et al., 2015)	√		MIAS	96%
(Rouhi et al., 2015)	√		MIAS DDSM	90.16% 96.47%
(Pak et al., 2015)	√		MIAS	91.43%
(Dong et al., 2015)	√		DDSM	97.73 %
(Sakikala and Ezhilarassi, 2016)	√		DDSM	92.6%
			In Breast	89.1%
(Valarmathie et al., 2016)	√		MIAS	98%
(Arevalo et al., 2016)		√	BCDR F03	82.6%
(Singh and Singh, 2016)	√		DDSM	75%
(Jiao et al., 2016)		√	DDSM	96.7%
(Rabidas et al ., 2016)	√		DDSM	92.25
(Abdel-Zaher et Eldeib, 2016)		√	Wisconsin	99.68%
(Saraswathi and Srinivasan, 2016)	√		MIAS	98.18%
(Suzuki et al., 2016)		√	DDSM	89.9%.
(Khan et al., 2016)	√		DDSM	93.95%

Chapitre 03 : Synthèse des travaux de classification des lésions mammaire et de super-résolution appliquée aux images médicales

(Jadoon et al., 2017)		√	IRMA	83,74%.
(Seryasat and Haddadnia, 2017)	√		MIAS DDSM	93% 90%
(Duraisamy and Emperumal, 2017)		√	MIAS	99%
(Prasad and Basha, 2018)	√		MIAS	92%
(Berbar, 2018)	√		MIAS DDSM	97.89% 98.69%
(Sheba and Raj, 2018)	√		MIAS	96%
(Braz et al., 2018)	√		MIAS DDSM	97.30% 91.63%
(Ribli et al., 2018)		√	DREAM	85%
(Pezeshki et al., 2019)	√		MIAS DDSM	91.37% 93.22%

Tableau 3.1 : Résumé des travaux de classification des lésions mammaires.

3.4 Application de la super-résolution sur les images médicales

Rappelons que la super-résolution (SR) est une technique qui reconstruit une image à plus haute résolution (HR) d'après une séquence d'images à basse résolution (BR), qui sont enregistrées dans une position commune et ensuite superposées afin de récupérer les informations sous-pixel, comme l'illustre la figure 3.5, ces informations supplémentaires permettent la reconstruction d'une image à haute résolution par la combinaison de différentes matrices associées à différentes vues du même objet (Wallach et al., 2012).

Diverses modalités d'imagerie médicale peuvent fournir à la fois des informations anatomiques sur la structure du corps humain et des informations fonctionnelles sur la santé (Yue et al., 2016) ; L'image médicale à haute résolution est la clé de la détection précoce des anomalies ou des pathologies (Pham, 2018), elle peut aussi accroître la précision de

l'évaluation de la taille et de la morphologie des organes et des pathologies (Greenspan, 2009) , cependant, les limites de résolution dégradent toujours la valeur des images médicales dans le diagnostic (Yue et al., 2016) ; en pratique, la fréquence d'échantillonnage maximale des détecteurs des appareils d'imagerie limite la portée des radiofréquences captées par l'objet imagé ; une solution pour augmenter la résolution consiste à réduire la taille des détecteurs, mais cela augmente le bruit et réduit donc le rapport signal-bruit. L'augmentation de la résolution des images par la technique de super-résolution est une clé pour une meilleure compréhension de l'anatomie (Greenspan, 2009 ; Pham, 2018) et pour préserver également la vérité (Yue et al., 2016).

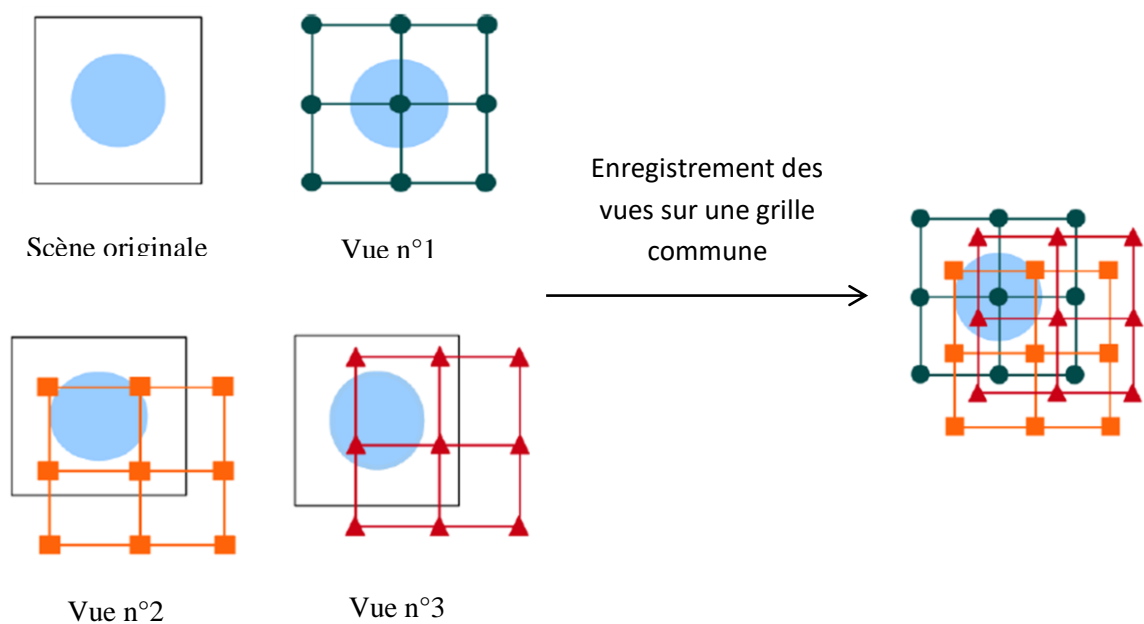


Figure 3.5 : Représentation graphique du processus de la super-résolution (Wallach et al., 2012).

Ces dernières années, plusieurs groupes de recherche ont commencé à aborder l'objectif d'augmentation de la résolution de l'imagerie médicale. La motivation de cette initiative est apparue à la suite d'avancées majeures dans les domaines du traitement de l'image et de la vidéo qui ont indiqué la possibilité d'augmenter la résolution en utilisant les algorithmes de super-résolution (Greenspan, 2009).

La super-résolution a été utilisée avec différents modalités d'imagerie médicale, notamment l'imagerie par résonance magnétique (IRM), l'IRM fonctionnelle (IRMf), la tomographie par émission de positrons PET (*positron emission tomography*) (Yue et al., 2016), et la mammographie ; la super-résolution des images médicales peut être utilisée pour améliorer les performances des méthodes de segmentation et d'enregistrement des images ; une

meilleure qualité d'image peut résulter en une segmentation et un enregistrement plus précis (Pham, 2018).

Divers algorithmes de super-résolution ont été proposés dans la littérature, où l'objectif commun est d'estimer précisément l'image à haute résolution, tout en minimisant le bruit et en préservant les contraintes importantes de l'image, notamment la fluidité de l'image et les connaissances préalables supplémentaires sur la source (Greenspan, 2009).

Kennedy et al. (Kennedy et al., 2006 dans Wallach, 2011) sont parmi les premiers qui ont appliqué la méthode de la super-résolution sur les images PET, où la séquence d'images a été créée en faisant subir des rotations et des translations à l'objet au cours de l'acquisition. Ensuite, ils ont reconstitué des données correspondant à chaque position, et ils ont appliqué la technique de la super-résolution à la séquence obtenue (Wallach, 2011). Leur méthode a été également appliquée à la fusion des images PET et TDM (tomodensitomètre) sur des données fantômes et cliniques (Kennedy et al., 2007 dans Wallach, 2011), où ils ont prouvé que les lésions mesurées sur les images à haute résolution avaient des tailles plus précises que celles de l'image acquise où il n'y a pas de mouvement, et un contraste plus élevé. La fusion de l'image à haute résolution à l'image TDM a augmenté le contraste par rapport à l'image elle-même ; la figure 3.6 illustre un exemple des images comparées dans leurs étude (Wallach, 2011).

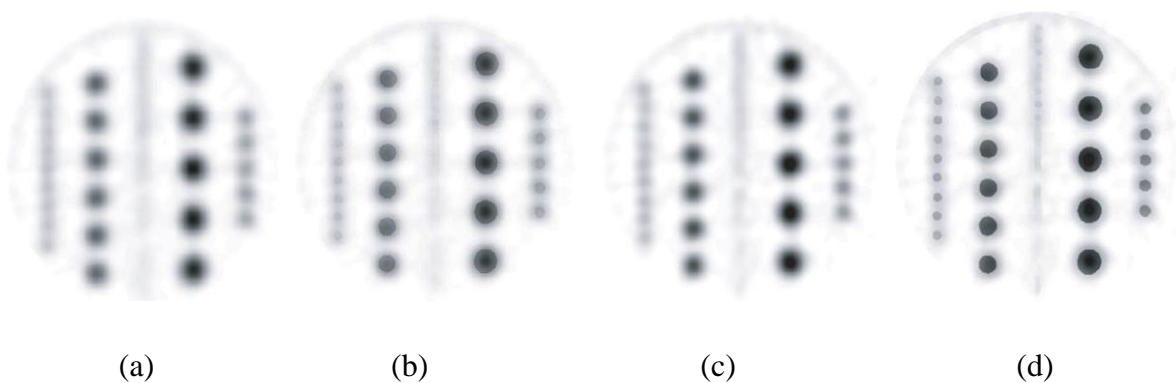


Figure 3.6 : (a) image acquise sans mouvement, (b) image acquise sans mouvement et fusionnée à l'image TDM, (c) image acquise avec des mouvements de rotation et de translation et à haute résolution, (d) image acquise avec des mouvements de rotation et de translation, à haute résolution et fusionnée à l'image TDM (Wallach, 2011).

Zheng et al. (Zheng et al., 2010) ont présenté une méthode qui permet de synthétiser des images mammographiques à haute résolution à partir des images à faible résolution, ce qui

offre la possibilité de faire un diagnostic précis tout en minimisant les risques pour les patientes. Leur algorithme combine des méthodes d'apprentissage statistiques et de recherche stochastique afin d'apprendre la cartographie (*mapping*) des mammographies, de la basse résolution à la haute résolution, en utilisant un grand ensemble de données de paires d'images d'entraînement, qui sont tirées de la base de donnée DDSM. Leurs résultats expérimentaux ont montré que l'algorithme de super-résolution peut générer des mammographies mammaires de haute qualité et de haute résolution à partir des données à basse résolution sans l'intervention humaine. La figure 3.7 illustre un exemple d'application de la technique de super-résolution sur une image mammographique.

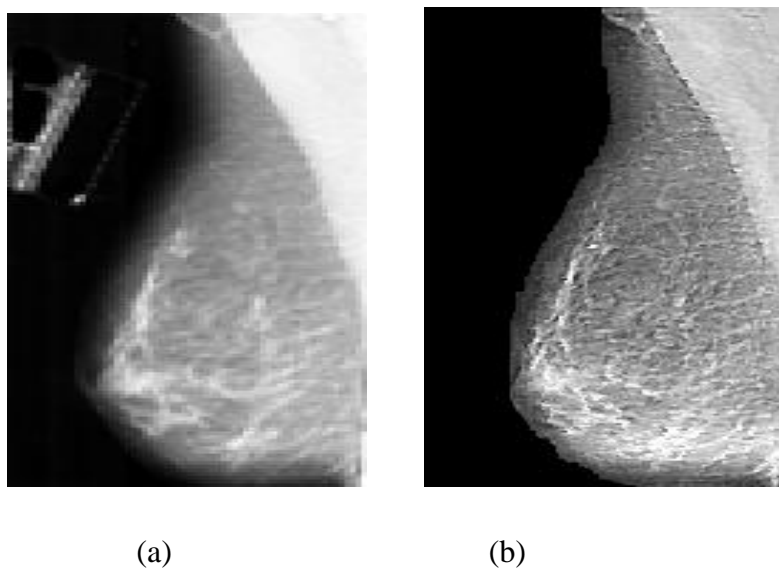


Figure 3.7 : Exemple d'application de la super-résolution sur une image mammographique (a : image à basse résolution ; b : image avec la super-résolution) (Zheng et al., 2010)

Wallach et al., 2012 (Wallach et al., 2012) ont mis en œuvre un algorithme de super-résolution par Maximum A Posteriori MAP (*a maximum a posteriori*) pour la compensation du mouvement respiratoire des images PET.

Afin de préserver les bords et assurer la convergence, une fonction de régularisation de Huber a été utilisée ; ils ont basé sur la fonction « B-splines » pour modéliser les mouvements élastiques qui sont souvent utilisés dans le cadre du mouvement respiratoire. Ils ont évalué la performance de l'algorithme proposé par l'utilisation des images simulées et cliniques en évaluant le faible rapport signal/bruit SNR (*low signal-to-noise ratio*) des images, ainsi que le contraste, la position et l'étendue des différentes lésions.

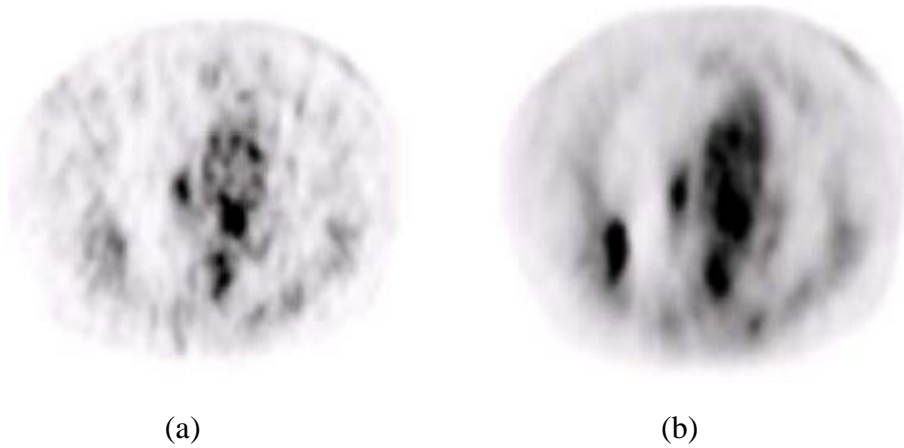


Figure 3.8 : Exemple d'application de la super-résolution sur l'image PET (a : image originale à basse résolution ; b : image à haute résolution) (Wallach and al., 2012)

La figure 3.8 illustre un exemple d'application de la super-résolution sur une image PET, où l'image à haute résolution avait un SNR (en moyenne plus de 4 fois plus élevé) et un contraste de la lésion 2 fois plus élevé que l'image synchronisée respiratoire unique.

L'utilisation de la technique de super-résolution appliquées aux images synchronisées avec le mouvement respiratoire a permis de compenser le mouvement avec une amélioration du SNR et du contraste de l'image, sans augmenter le temps d'acquisition global (Wallach and al., 2012).

Trinh et al., 2014 (Trinh et al., 2014) ont proposé une méthode appelée « super-résolution par poids faible » SRSW (*Super-Resolution by Sparse Weight*) pour le débruitage et la super-résolution des images médicales. Le débruitage et la super-résolution sont effectués sur chaque patch (image) d'entrée à basse résolution, et sa version à haute résolution est estimée en trouvant une représentation linéaire non négative du patch d'entrée par rapport aux patches à basse résolution de la base de données, où les coefficients de la représentation dépendent fortement de la similarité entre le patch d'entrée et les patches échantillons de la base de données. Leur méthode proposée est particulièrement utile dans les cas du bruit corrompu et des images à basse résolution. Ils ont confirmé que la méthode proposée est performante, car elle élimine efficacement le bruit. Les figures 3.9 et 3.10 illustrent deux images IRM avant et après l'application de la super-résolution.

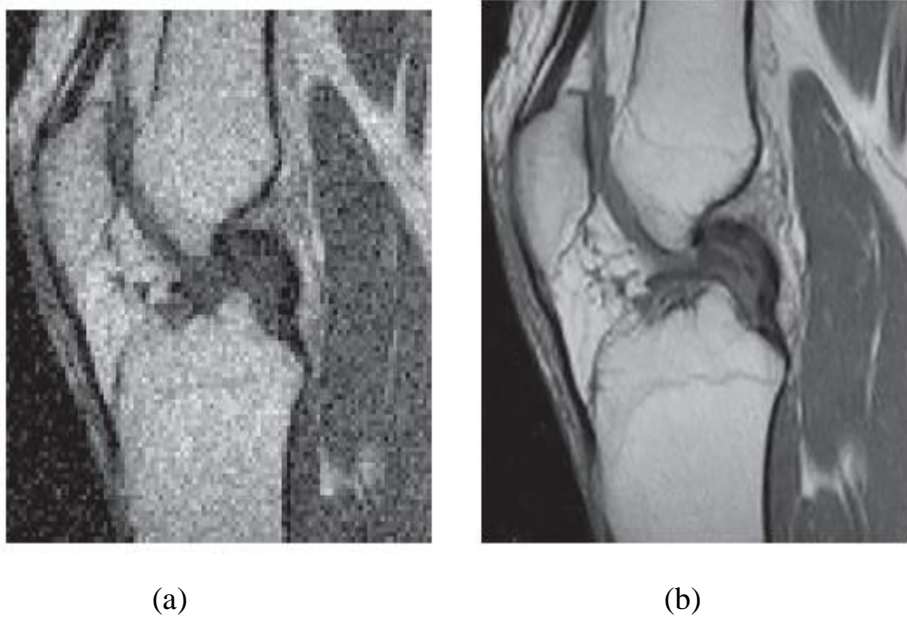


Figure 3.9 : Image IRM du genou (a : image originale à BR ; b : image à HR) (Trinh et al., 2014)

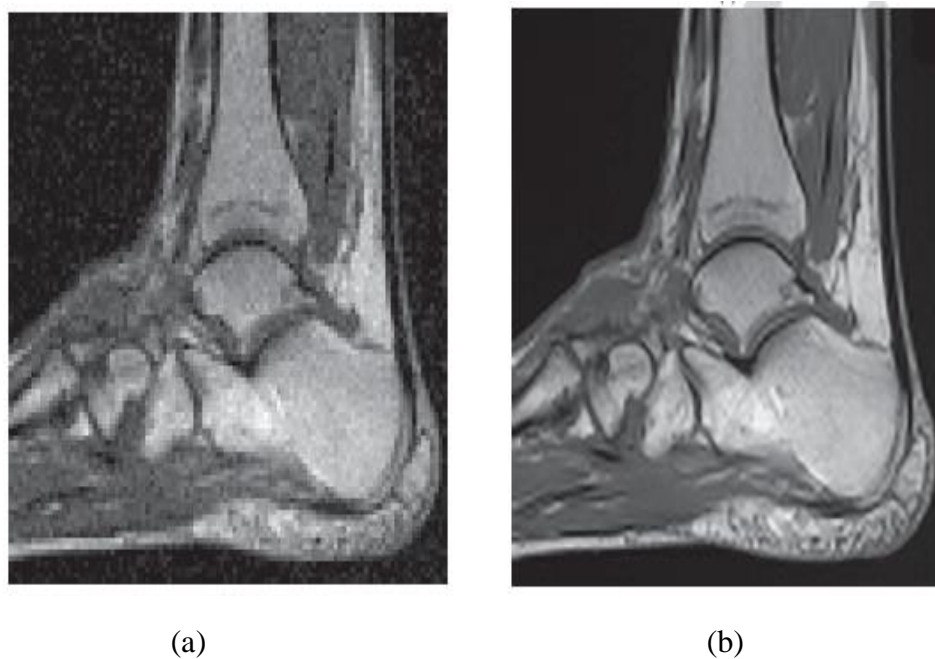


Figure 3.10 : Image IRM de la cheville (a : image originale à BR ; b : image à HR) (Trinh et al., 2014)

Afin de récupérer une image à haute résolution à partir d'une seule image à basse résolution, Shi et al. (Shi et al. 2015) ont proposé une nouvelle méthode de super-résolution appelée méthode de variation totale de faible rang LRTV (*Low-Rank Total-Variation*), ils ont combiné des régularisations de bas rang et de variation totale, cette combinaison a permis de

rassembler des informations globales et locales pour une récupération efficace de l'image à haute dimension ; leur méthode est appliquée sur les images du cerveau, et ils ont montré qu'elle préserve au mieux les arêtes, elle permet également d'obtenir les valeurs SNR élevées, et elle est capable de remédier l'effet du volume partiel et de récupérer les détails de la structure fine du cerveau à partir d'images adultes et pédiatriques. La figure 3.11 illustre une image de cerveau d'un adulte avant et après l'application de la super-résolution, où la valeur du SNR = 24.9.

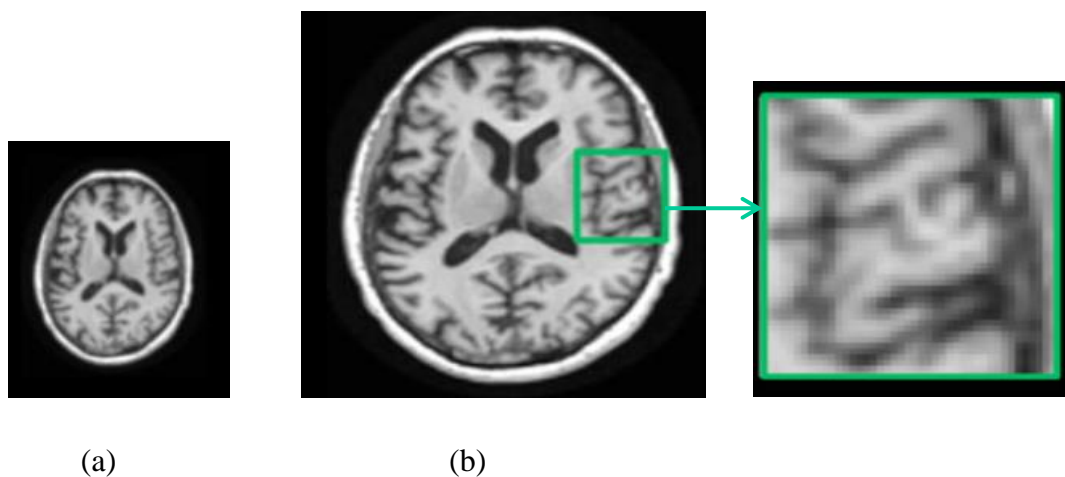


Figure 3.11 : Résultats du sur-échantillonnage d'une image d'un adulte (a : image originale, b : image à haute résolution) (Shi et al. 2015).

Odille et al. (Odille et al., 2015) ont utilisé une méthode pour augmenter la résolution des images 3D ciné-IRM du cœur. Pendant la respiration libre, une séquence d'images 2D (*stacks*) avec différentes orientations sont acquises ; la reconstruction avec la super-résolution a été mise en œuvre avec deux méthodes de régularisation, la première appelée (*Tikhonov*) qui est une méthode conventionnelle, et la deuxième appelée (*Beltrami*) est une méthode qui préserve les caractéristiques.

Des exemples d'images reconstruites sont présentés dans la figure 3.12, où les images natives sont rangées en haut, les images à haute résolution avec la régularisation de Tikhonov sont rangées au milieu, et les images à haute résolution avec la régularisation de Beltrami sont rangées en bas ; les images sont montrées en position 'fin-diastole' et en position 'fin-systole' ; les flèches indiquent les détails anatomiques notables qui n'étaient pas visibles sur les images natives, mais qui ont été récupérés par les deux techniques de super-résolution, en particulier près de la base et de l'apex du ventricule gauche (Odille et al., 2015).

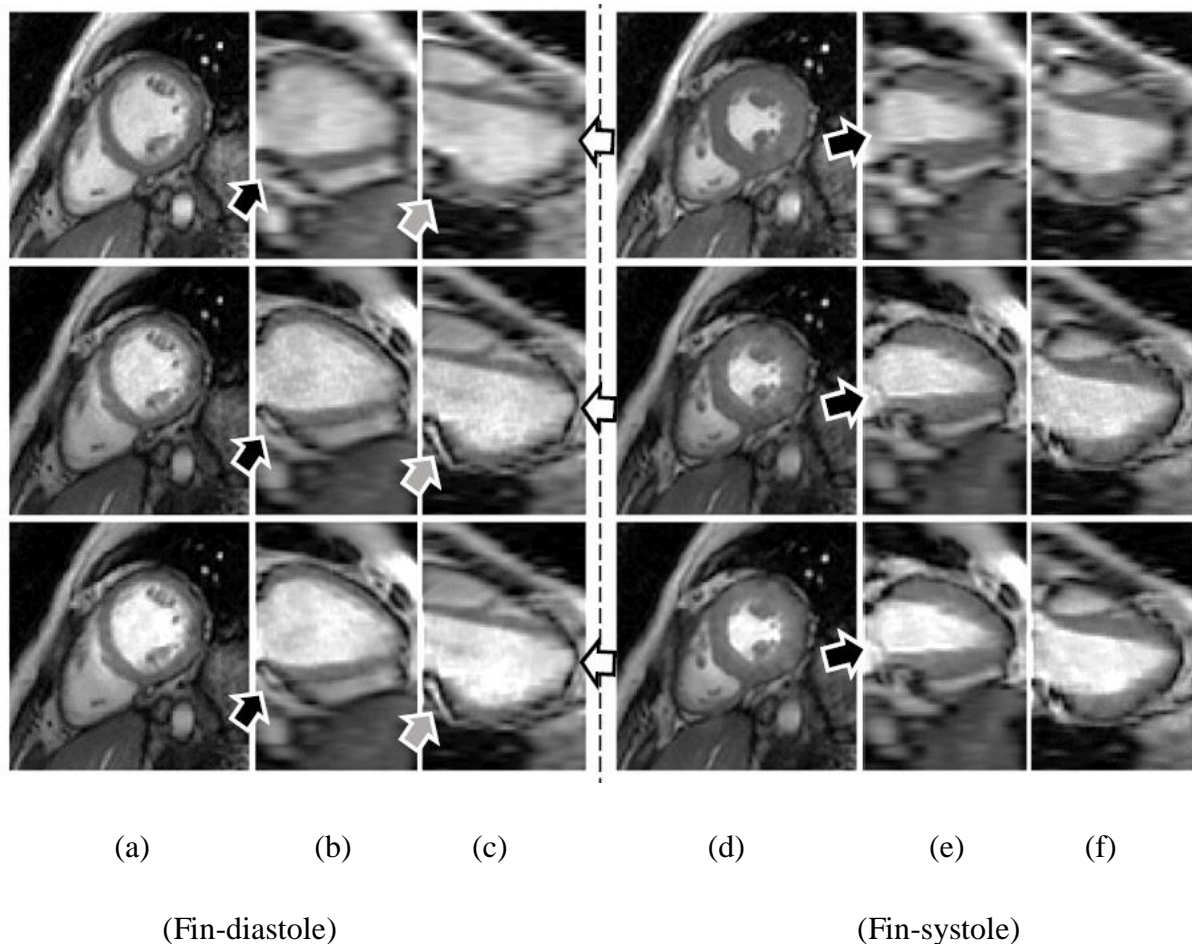


Figure 3.12 : Exemple d'images reconstruites où : (a) et (d) sont des orientations à court axe, (b) et (e) sont des orientations à long axe vertical, et (c) et (f) sont des orientations à long axe horizontal (Odille et al., 2015).

Afin de construire des images mammographiques à haute résolution à partir des images à basse résolution, Umehara et al. (Umehara et al., 2017) ont utilisé une méthode de super résolution basée sur l'apprentissage profond appelée SRCNN (*super-resolution convolutional neural network*).

Leurs résultats obtenus avec les mammographies cliniques ont révélé que la méthode SRCNN peut améliorer considérablement la qualité de l'image des mammographies, en particulier dans les cas des seins denses, des groupes d'évaluation BI-RADS à haut risque et des cas malins vérifiés par pathologie. La figure 3.13 montre un exemple d'image obtenue à l'aide du schéma SRCNN dans un cas malin.

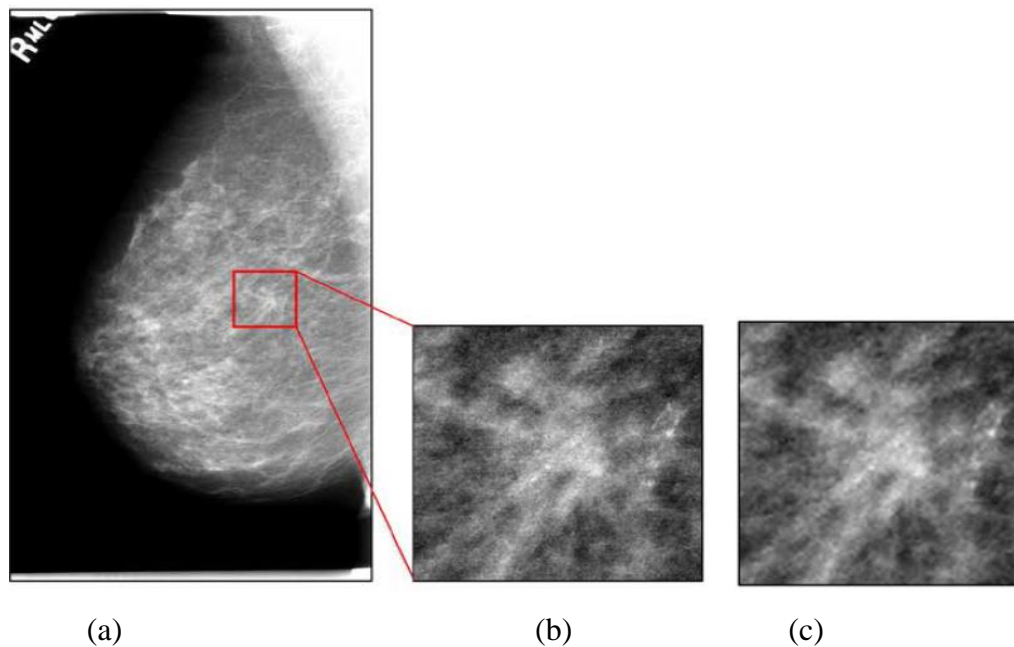


Figure 3.13 : Exemple d'image reconstituée à l'aide de la super-résolution (a : image oblique médio-latérale originale, b : image originale de la région d'intérêt, c : image avec SRCNN (Umehara et al., 2017)).

Les auteurs Greenspan (Greenspan, 2009), Christensen-Jeffries et al. (Christensen-Jeffries et al., 2020) et Yue et al. (Yue et al., 2016) ont présenté un survol détaillé sur les travaux qui ont basé sur la méthode de super résolution dans le domaine médical, où Greenspan (Greenspan, 2009) a présenté un aperçu sur l'intégration de la technique de super-résolution dans les applications d'imagerie médicale, il a présenté également les défis liés à la résolution dans ce domaine ; il a donné un bref aperçu des algorithmes de super-résolution, et il a concentré sur l'application de la super-résolution sur les images IRM et PET.

Christensen-Jeffries et al. (Christensen-Jeffries et al., 2020) ont également présenté un survol des travaux récents basés sur la technique de la super résolution, et qui sont appliqués essentiellement sur les images échographiques.

Yue et al. (Yue et al., 2016) ont présenté les principales contributions basées sur la super-résolution dans ces dernières années dans plusieurs domaines (surveillance, diagnostic médical, télédétection pour l'observation de la terre ,etc.), ils ont également détaillé les principaux défis de la super-résolution d'image.

Les travaux qui sont basés sur la super-résolution ont prouvé l'importance et l'efficacité de cette technique dans le domaine médicale ; malgré les défis liés à cette méthode, elle est

toujours capable d'améliorer la qualité des images médicale, ce qui aide les spécialistes à faire un bon diagnostic.

3.5 Conclusion

Dans ce chapitre, nous avons présenté un survol sur les notions clé de notre thèse à savoir la caractérisation et la classification des masses mammaires et l'utilisation de la technique de la super résolution dans le domaine médical.

Dans un premier temps, une grande attention est portée aux méthodes d'extraction de caractéristiques et de classification des lésions mammaires dans les images mammographiques. Ces méthodes se distinguent par plusieurs points comme le type du descripteur, la taille de la base de données, le type de la lésion, l'approche de classification. D'après tous les travaux cités, nous pouvons confirmer que les caractéristiques de texture jouent un rôle important pour différencier la masse maligne de la masse bénigne, c'est pourquoi, dans toutes nos approches proposées, nous allons utiliser ce type de caractéristique afin d'améliorer les performances de la classification.

Dans un deuxième temps, nous avons synthétisé les travaux qui sont basés sur la super-résolution dans le domaine médical, et nous avons confirmé que cette méthode a été appliquée avec succès, car elle augmente considérablement la qualité des images médicales quelque soit la modalité d'image utilisée, pour cette raison, nous allons l'utiliser dans l'étape de discrimination.

Dans la prochaine partie de cette thèse, nous décrirons nos contributions proposées.

Partie II :

Contributions

Chapitre 4 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales.

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture.

Chapitre 4

Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

4.1 Introduction

Dans le présent chapitre, nous présentons dans un premier temps toutes les ressources que nous avons utilisées dans nos travaux. Dans un deuxième temps, nous abordons la conception et l'étude expérimentale des différentes étapes constituant la première contribution proposée qui comporte quatre expérimentations, et qui a pour objectif la classification des régions d'intérêt dans les images mammographiques normales, et qui est basée principalement sur les attributs de texture. Enfin, nous synthétisons les résultats atteints au niveau de chaque expérimentation.

Nous montrerons au cours de ce chapitre, et à travers plusieurs expérimentations, que l'utilisation des attributs de texture donne des résultats satisfaisants du point de vue de la classification. Nous aborderons dans les détails les quatre différentes expérimentations :

- **1ère expérimentation** : Hybridation des caractéristiques statistiques.
- **2ème expérimentation** : Hybridation des descripteurs du premier ordre et de Haralick.
- **3ème expérimentation** : Classification des masses mammaires basée sur les attributs statistiques du premier ordre.
- **4ème expérimentation** : Représentation profonde de la texture via l'opérateur LBP.

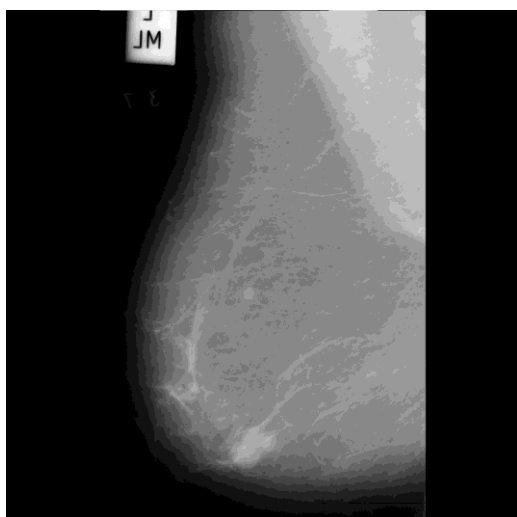
4.2 Ressources utilisées

4.2.1 Ordinateur

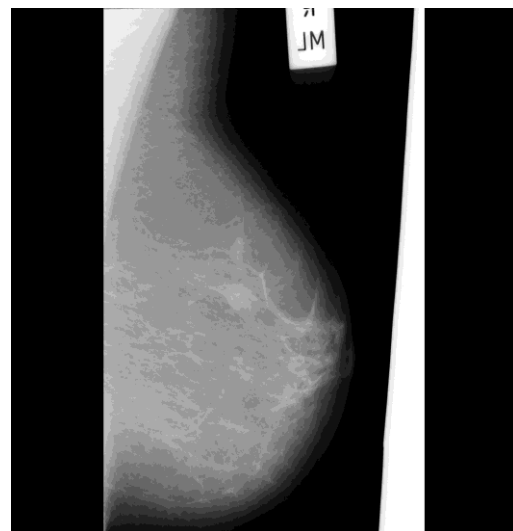
L'intégralité des expérimentations décrites dans cette thèse a été réalisée sur un ordinateur portable de type ASUS, où le système d'exploitation : Windows 10 famille, 64 bits ; le processeur : Intel ® Core ™, 7-4720 HQ CPU @ 2.60 GHZ ; Mémoire vive : 8.00 Go.

4.2.2 Base des images mammographiques

Lors de cette étude, nous avons utilisé la base des images MIAS (*Mammographic Image Analysis Society*) qui est une organisation de groupes de recherche britanniques qui s'intéressent à la compréhension et à la création d'une base de données des images mammographiques numériques. Cette base de données contient 322 films numérisés et elle comprend des marques de "vérité" du radiologue sur les emplacements de toute anomalie qui pourrait être présente. La base de données a été réduite à un bord de 200 microns de pixels et coupée de telles sortes que toutes les images soient en 1024x1024 (web 9). La figure 4.1 représente des échantillons de la base de données MIAS.



(a)



(b)

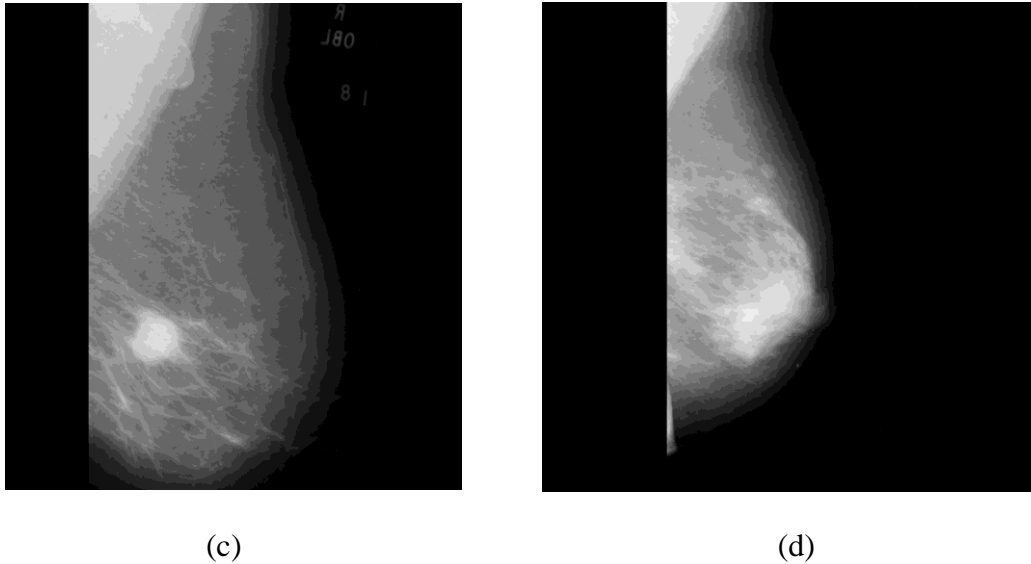


Figure 4.1 : Echantillons de la base de données MIAS (a : image bénigne n° 5, b : image bénigne n° 12, c : image maligne n° 28, d : image maligne n° 72)

4.2.3 Environnement de travail

4.2.3.1 Eclipse

Eclipse est un IDE (*Integrated Development Environment*), c'est-à-dire un logiciel qui simplifie la programmation avec un certain nombre de raccourcis qui aide à la programmation. Il est développé par IBM, il est gratuit et disponible via le site (<http://www.eclipse.org/downloads>) pour la plupart des systèmes d'exploitation (web 10).

4.2.3.2 NetBeans

NetBeans est un IDE (outil de développement intégré pour Java). Il est possible d'étendre le support à d'autres langages de programmation à l'aide d'extension (pour supporter par exemple le PHP, C, C++, HTML, JavaScript) (web 11).

4.2.3.3 Java

En ce qui concerne le langage de programmation, nous avons opté pour Java qui est un langage de programmation à usage général, évolué et orienté objet dont la syntaxe est proche du C. Ses caractéristiques ainsi que la richesse de son écosystème et de sa communauté lui ont permis d'être très largement utilisé pour le développement d'applications. Java est notamment largement utilisé pour le développement d'applications d'entreprises et mobiles. Parmi les caractéristiques du langage Java c'est qu'il est simple, fortement typé, interprété, et orienté objet (web 12).

4.2.3.4 Weka

WEKA (*Waikato Environment for Knowledge Analysis*) qui est un outil de data-mining, open source et il est basé sur Java. WEKA est un ensemble de nombreux algorithmes d'exploration de données et d'apprentissage automatique, comprenant le prétraitement des données, la classification, le regroupement et l'extraction des règles d'association (voir figure 4.2). Ces algorithmes peuvent être appliqués directement à un ensemble de données ou peuvent être appelés à partir d'un code Java (Rangra et Bansal 2014).

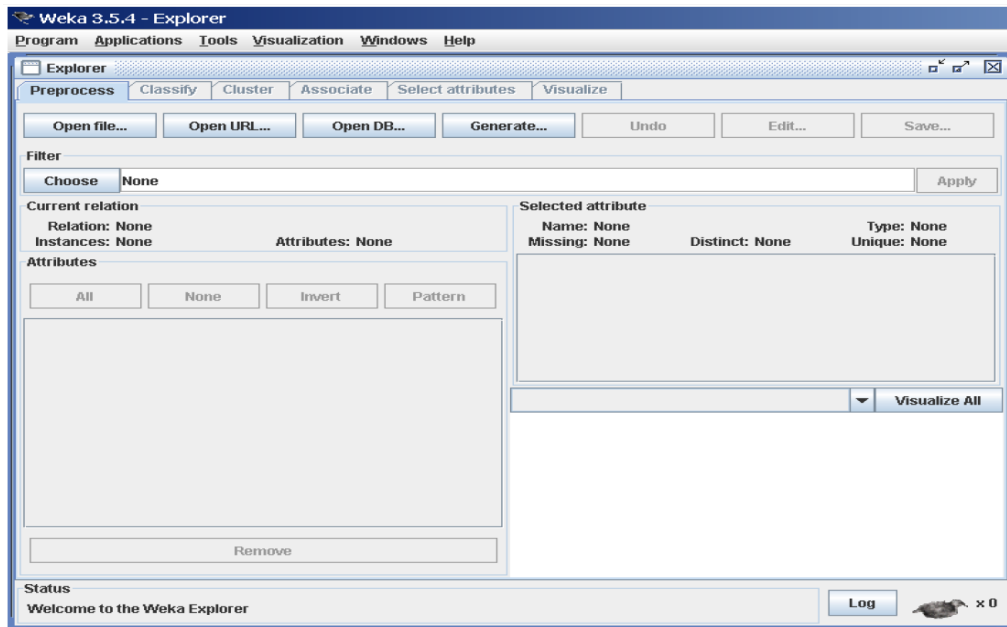


Figure 4.2 : Exploration des options de Weka

4.2.3.5 ImageJ

ImageJ est un logiciel de traitement d'images et d'analyse d'image scientifique, qui est riche d'un ensemble de fonctions délivrées sous forme de plugins. Derrière une interface qui est un peu désuète, ImageJ cache des fonctions simples d'accès, directes et puissantes pour l'analyse d'image (Sage, 2012). ImageJ est disponible sur le site (<http://rsb.info.nih.gov/ij/download.html>).

4.2.4 Critères d'évaluation

Pour évaluer les méthodes proposées, nous nous sommes basés sur le calcul de trois mesures importantes : la précision, la sensibilité et la spécificité.

4.2.4.1 La précision

La précision d'un test est sa capacité de différencier correctement les cas de tumeurs malignes et bénignes. Pour estimer la précision d'un test, nous devons calculer la proportion de vrais positifs et de vrais négatifs dans tous les cas évalués. Mathématiquement, cela est exprimé comme suit (Baratloo et al., 2015):

$$\text{La précision} = \frac{VP + VN}{VP + VN + FP + FN}$$

4.2.4.2 La sensibilité

La sensibilité d'un test est sa capacité de déterminer correctement les cas de malins. Pour l'estimer, il faut calculer la proportion de vrais positifs dans les cas malins. Mathématiquement, cela est exprimé comme suit (Baratloo et al., 2015):

$$\text{La sensibilité} = \frac{VP}{VP + FN}$$

4.2.4.3 La spécificité

La spécificité d'un test est sa capacité de déterminer correctement les cas bénins. Pour l'estimer, il faut calculer la proportion de vrais négatifs dans les cas bénins. Mathématiquement, cela est exprimé comme suit (Baratloo et al., 2015):

$$\text{La spécificité} = \frac{VN}{VN + FP}$$

Où :

Vrai positif (VP) = le nombre de cas correctement classés comme masse maligne.

Faux positif (FP) = le nombre de cas incorrectement classés comme masse maligne.

Vrai négatif (VN) = le nombre de cas correctement classifiés comme masse bénigne.

Faux négatif (FN) = le nombre de cas incorrectement classés comme masse bénigne.

4.2.4.4 La matrice de confusion

Une matrice de confusion (MC) est un tableau qui est souvent utilisé pour décrire la performance d'un modèle de classification (ou classifieur) sur un ensemble de données d'essai dont les valeurs réelles sont connues. La matrice de confusion elle-même est relativement

simple à comprendre, mais la terminologie qui s'y rapporte peut prêter à confusion (web 13). Le tableau 4.1 représente la matrice de confusion.

Classe	Correcte	
	Masse maligne	Masse bénigne
Masse maligne	Vrai positif (VP)	Faux positif (FP)
Masse bénigne	Faux négatif (FN)	Vrai négatif (VN)

Tableau 4.1 : La matrice de confusion

4.2.4.5 L'aire sous la courbe ROC

La courbe ROC (Receiver Operating Characteristic) est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test pour toutes les valeurs seuils possibles. L'ordonnée représente la sensibilité et l'abscisse correspond à la quantité (1 - spécificité) (Delacour et al., 2005).

Il est important de noter plusieurs points dans l'espace ROC, le point inférieur gauche (0, 0) représente la stratégie consistant à ne jamais émettre une classification positive. La stratégie opposée, qui consiste à émettre sans condition des classifications positives, est représentée par le point supérieur droit (1, 1). Le point (0, 1) représente la classification parfaite.

Pour comparer les classificateurs, la performance ROC doit être réduite à une seule valeur scalaire représentant la performance attendue. Une méthode courante consiste à calculer l'aire sous la courbe ROC (AUC). Comme AUC est une partie de l'aire de l'unité de mesure carrée, sa valeur sera toujours comprise entre 0 et 1. Cependant, comme la supposition aléatoire produit la diagonale entre (0, 0) et (1, 1), qui a une aire de 0,5, aucun classificateur réaliste ne devrait avoir une AUC inférieure à 0,5 (Fawcett, 2006).

4.3 Extraction de l'information de texture par les descripteurs statistiques du premier et du deuxième ordre

Déterminer si une masse mammaire est bénigne ou maligne n'est pas une tâche simple. Dans certains cas les deux types de masses présentent des similarités que même un expert a du mal à les différencier. Les systèmes d'aide au diagnostic peuvent apporter une aide dans ce sens. Une de leurs tâches consiste à extraire les caractéristiques pertinentes et à les utiliser dans des méthodes de classification pour déterminer le type de la masse. La réussite de ces systèmes

est étroitement liée au choix des caractéristiques. Dans tous nos travaux que nous allons présenter dans cette section, nous allons chercher quelle est la bonne combinaison de caractéristiques, nous allons également tester l'efficacité de plusieurs techniques de classification sur les formules de caractéristiques choisis, pour décider à la fin si les masses sont de type bénin (ne sont pas cancéreuses) ou de type malin (sont cancéreuses).

Dans cette section, nous allons proposer quatre systèmes, et dans chacun, nous avons utilisé l'outil ImageJ pour segmenter manuellement les régions d'intérêts en se basant sur la position du centre de masse ; Rappelons que l'étape de prétraitement et l'étape de segmentation ne font pas l'objet de notre travail, nous nous soucions de l'étape d'extraction des caractéristiques et l'étape de classification. En d'autres termes, on dispose d'une région masse, nous devons décider si elle est maligne ou non ; et puis nous utilisons l'outil Weka pour l'utilisation des classifieurs.

4.3.1 Hybridation des caractéristiques statistiques

Notre première idée clé est la fusion des caractéristiques de texture du premier et du second ordre en vue de répondre à la question suivante : la masse est-elle bénigne ou maligne ?

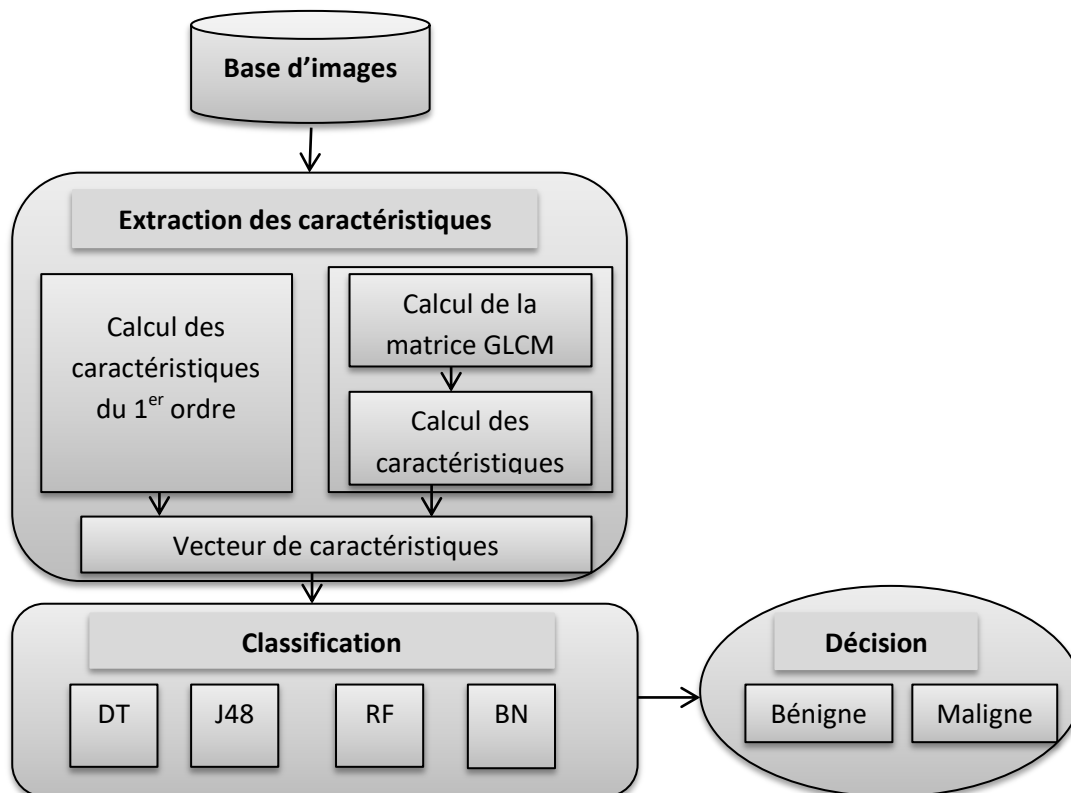


Figure 4.3 : Schéma général de l'approche proposée.

Notre objectif principal est de mettre en valeur l'efficacité de l'application de la notion de texture sur les masses mammaires. Ce premier travail a fait l'objet de la production scientifique de (Boudraa et al, 2017a), et les principales étapes de cette expérimentation sont récapitulées dans la figure 4.3.

4.3.1.1 Extraction des caractéristiques

L'extraction des caractéristiques est une méthode de capture du contenu visuel d'une image, l'objectif du processus d'extraction de caractéristiques est de représenter une image brute sous sa forme réduite pour faciliter la décision. Le bon choix des caractéristiques est l'étape clé qui permet d'obtenir un taux de classification élevé.

L'approche adoptée consiste à collecter douze caractéristiques de texture ; six caractéristiques du premier ordre et six caractéristiques du second ordre. Notons que plusieurs travaux ont confirmé l'importance de ces caractéristiques dans le domaine d'imagerie médicale. Les caractéristiques doivent pouvoir révéler les similitudes entre les objets d'une classe et en même temps, pouvoir montrer la différence entre eux et les autres objets des autres classes (Seryasat et Haddadnia 2017), c'est pourquoi notre travail se focalise sur la recherche d'une meilleure combinaison de caractéristiques de texture qui nous mènera à une classification de la masse (bénigne ou maligne) jugée satisfaisant.

Nous avons choisis six caractéristiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie et la régularité) qui ont été calculées à partir de la région d'intérêt complète, et six caractéristiques du second ordre qui ont été définies par Haralick et al. (Haralick et al., 1973). Les caractéristiques retenues dans l'approche proposée sont les suivantes : le moment angulaire du second degré, le contraste, la corrélation, la variance, le moment des différences inverses et l'entropie. Ces caractéristiques sont calculées à partir de la matrice de cooccurrences (GLCM) de l'image (appliquée à la région d'intérêt uniquement) ; nous nous sommes basés sur la moyenne des directions angulaires classiquement utilisées (0° , 45° , 90° et 135°) ; La figure 4.4 illustre un exemple d'une matrice GLCM calculée en utilisant 1 pour la distance entre les pixels et 0° pour la direction, appliquée à une image mammographique (Jinan, 2015).

Rappelons que c'est le groupe de Haralick et al. (Haralick et al., 1973) qui a proposé les premiers quatorze paramètres de texture calculés à partir de la matrice de cooccurrences. Le tableau 4.2 résume les caractéristiques sélectionnées avec leurs formules, et le tableau 4.3

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

illustre un exemple de deux vecteurs de caractéristiques de deux masses l'une maligne (n°115) et l'autre bénigne (n°17).

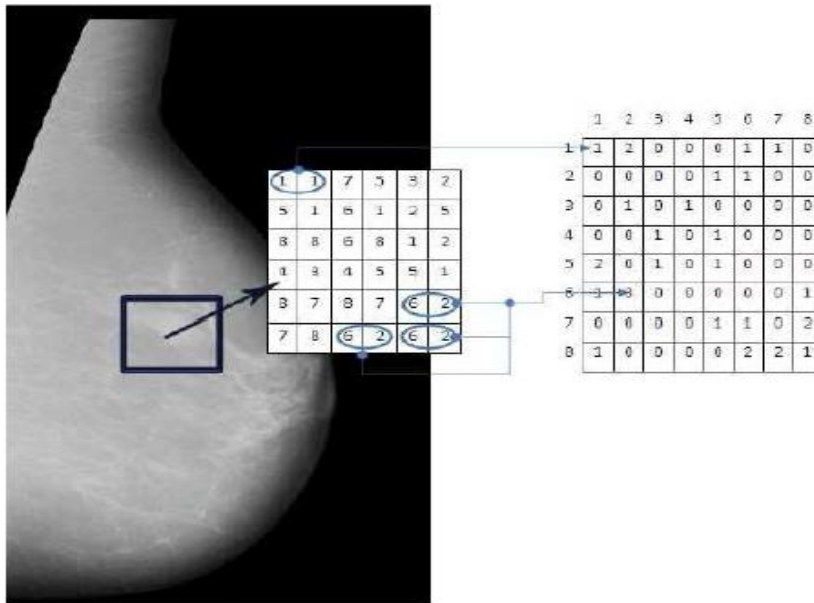


Figure 4.4 : Exemple d'une matrice GLCM (Jinan, 2015).

	Caractéristique	Formule
Caractéristiques du 1 ^{er} ordre	Moyenne	$\frac{1}{N} \sum_i \sum_j I(i, j)$
	Variance	$\sqrt{\frac{\sum_i \sum_j (P(i, j) - m)^2}{N}}$
	Asymétrie	$\frac{\sum_{i=0}^{n-1} P_i (x_i - m)^3}{\sum_{i=0}^{n-1} P_i (x_i - m) 3/2}$
	Aplatissement	$\frac{1}{N} \frac{\sum_i \sum_j (P(i, j) - m)^4}{Std^4}$
	Entropie	$-\sum_i \sum_{ij} P(i, j) \log P(i, j)$
	Régularité	$1 - \frac{1}{1 + \sigma^2(z)}$

Caractéristiques de Haralick	Moment angulaire du second degré (Energie)	$\sum_i \sum_j P_i(i, j)^2$
	Contraste	$\sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} P(i, j) \right\}_{ i-j =n}$
	Corrélation	$\frac{\sum_i \sum_j (ij) P_i(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
	Variance	$\sum_i \sum_j (i - \mu)^2 P(i, j)$
	Moment des différences inverses	$\sum_i \sum_j \frac{1}{1+(i-j)^2} P(i, j)$
	Entropie	$-\sum_i \sum_j P(i, j) \log(P(i, j))$

Tableau 4.2 : Caractéristiques du premier et du deuxième ordre

Type	Caractéristique	Masse maligne (image n° 115)	Masse bénigne (image n° 17)
Caractéristiques du 1 ^{er} ordre	Moyenne	213.90105988617972	188.8341684935797
	Variance	116.09123464988419	543.3698738741471
	Asymétrie	3.985194425638405	1.7863248627327062
	Aplatissement	-9.360293038121865	-20.668336987165024
	Entropie	0.9201672344384454	0.608680028745505
	Régularité	0.9150711670283276	0.9588651646732125
Caractéristiques de Haralick	Moment angulaire du second degré	0.7750874326746846	0.9203761081581778
	Contraste	138.30021289125065	97.05081625586521
	Corrélation	2.0309705554155322E-4	6.924152682044664E-4
	Variance	4853.604758866137	1393.9444503414047
	Moment des différences inverses	0.9568207554176146	0.9821884296922628

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

	Entropie	0.5840479352625975	0.22752495589524066
--	----------	--------------------	---------------------

Tableau 4.3 : Exemple de deux vecteurs de caractéristiques

4.3.1.2 Classification

Notons que dans toutes nos expérimentations, nous avons testé plusieurs techniques de classifications, et nous avons gardé que les classifieurs qui ont donné les meilleurs résultats.

Le processus de classification se déroule de la façon suivante : pour classifier une lésion, nous soumissions le vecteur de caractéristiques de cette dernière aux quatre méthodes de classification qui se trouvent dans WEKA (DT : *Decision Table*, J48, RF : *Random Forest* et BN : *Bayes- Network*) pour décider à la fin à quelle classe appartient cette lésion.

4.3.1.3 Résultats et discussion

Afin d'évaluer notre travail, nous nous sommes basés sur plusieurs métriques importants :

La précision (la capacité de différencier correctement les cas malins et les cas bénins), la sensibilité (la capacité de déterminer correctement les cas malins), la spécificité (la capacité de déterminer correctement les cas bénins), la matrice de confusion et la mesure AUC.

En analysant les résultats présentés dans le tableau 4.4, nous pouvons remarquer que la méthode de classification J48 a fourni le meilleur taux de détection qui égal 80.4 % avec un taux de sensibilité élevé de 91.6% et une spécificité de 70.3%, et la mesure AUC = 0.70 qui est un signe positif et proche de la valeur 1, c'est-à-dire que la méthode proposée est prometteuse.

Classifieur	MC		Précision	Sensibilité	Spécificité	AUC
DT	9	12	0.75	0.705	0.769	0.718
	5	30				
J48	10	11	0.804	0.916	0.772	0.703
	1	34				
RF	7	14	0.75	0.666	0.8	0.723
	7	28				
BN	6	15	0.786	0.714	0.828	0.753
	6	29				

Tableau 4.4 : Les métriques d'évaluation utilisées

Nous avons comparé notre travail avec la méthode qui utilise uniquement les descripteurs de Haralick (le moment angulaire du second degré, le contraste, la corrélation, la variance, le moment des différences inverses et l'entropie), notons que dans cette comparaison, nous avons utilisé le même nombre d'images (21 de type malin et 35 de type bénin), nous avons également utilisé les mêmes classifieurs.

Le tableau 4.5 illustre le pourcentage de détection correcte en utilisant notre combinaison de douze caractéristiques et celle qui se focalise uniquement sur les six caractéristiques de texture de Haralick et al. (Haralick et al., 1973). Nous apercevons clairement dans le tableau ci-dessous que l'approche proposée a permis d'obtenir de bons résultats (quel que soit le classifieur utilisé), et aussi par rapport aux résultats obtenus par l'approche basée uniquement sur les caractéristiques de Haralick et al. (Haralick et al., 1973).

Classifieur Méthode	Méthode proposée	Méthode basée sur les caractéristiques de Haralick et al. (Haralick et al.,1973)
DT	75%	55.35%
J48	80.36%	48.21%
RF	75%	62.5%
BN	78.57%	55.35%

Tableau 4.5 : Résultats de classification obtenus par la méthode proposée et par la méthode de Haralick (Haralick et al., 1973).

4.3.2 Hybridation des descripteurs du premier ordre et de Haralick

Dans le travail précédent, nous avons utilisé uniquement douze caractéristiques statistiques, et les résultats obtenus sont jugés satisfaisants ; notre objectif reste le même c'est-à-dire trouver la bonne combinaison des descripteurs qui donne à la fin un taux de classification élevé, c'est pourquoi, nous essayons de rajouter à l'ancienne combinaison deux autres caractéristiques statistiques et les évaluer par plusieurs techniques de classification, Ce travail a fait l'objet de la production scientifique de (Boudraa et al., 2017b) .

La figure 4.5 résume le processus général du notre méthode.

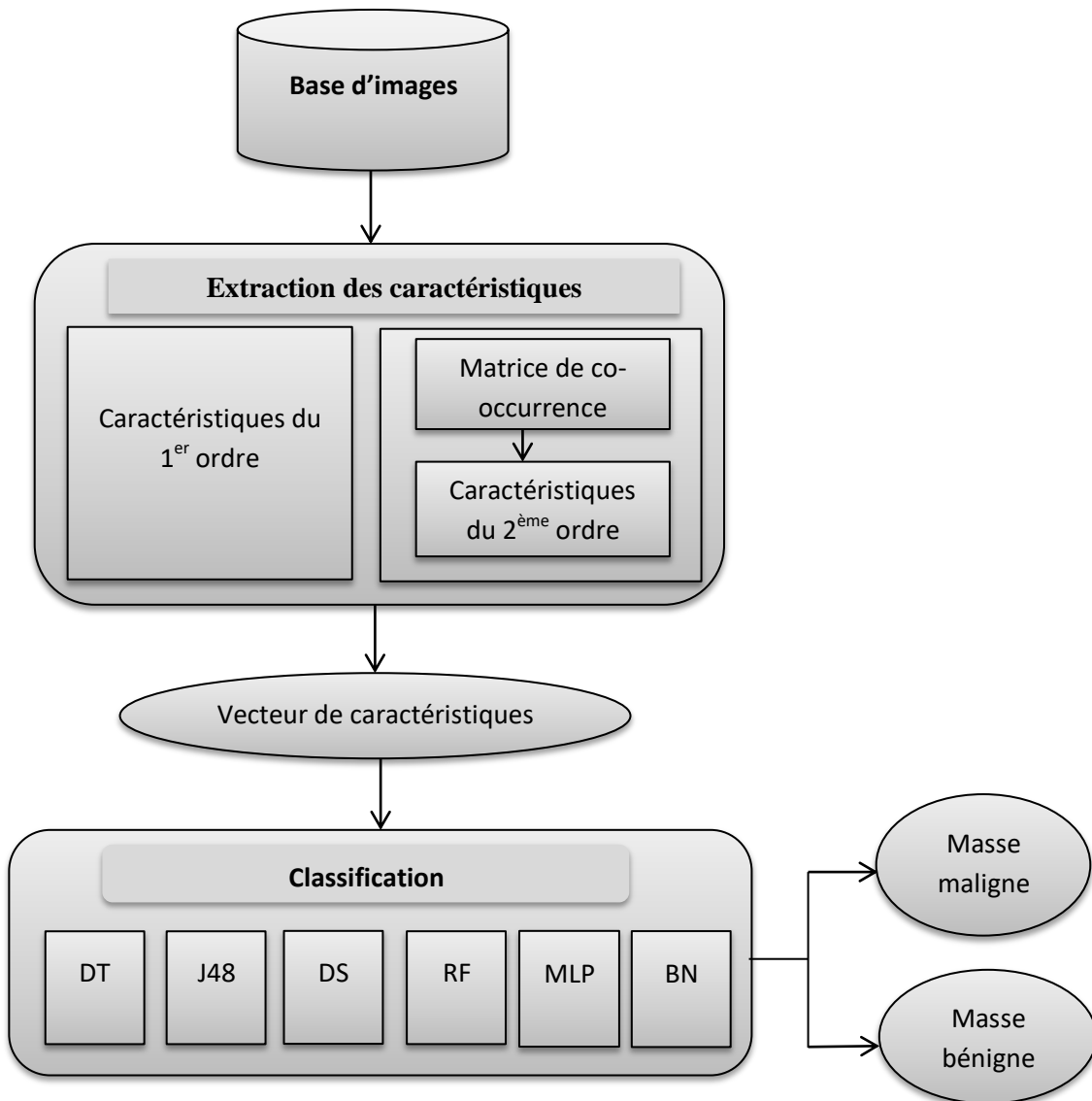


Figure 4.5 : Le processus de la méthode proposée

4.3.2.1 Extraction des caractéristiques

Comme nous avons déjà mentionné précédemment, plusieurs travaux dans la littérature ont montré l'importance des caractéristiques statistiques du premier ordre et les caractéristiques du GLCM qui peuvent distinguer les deux types de masses (maligne et bénigne) dans les images mammaires. Dans ce travail, plusieurs caractéristiques de texture du premier ordre et du deuxième ordre ont été rassemblées afin de déterminer la malignité d'une masse mammaire.

Dans un premier temps, nous calculons les caractéristiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'uniformité et la régularité) à partir de l'image complète, et nous calculons également la matrice de co-occurrences de chaque région d'intérêt, ainsi que le calcul des différents descripteurs du deuxième ordre à partir des

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

matrices GLCMs obtenues (le moment angulaire du second degré, le contraste, la corrélation, la variance, le moment des différences inverses, l'entropie, et CP (Cluster Prominence)).

Dans un deuxième temps, nous fusionnons toutes les caractéristiques calculées dans un vecteur de caractéristiques, et nous le soumissionnons aux différents classifieurs.

Le tableau 4.6 décrit les formules mathématiques de l'uniformité et le Cluster Prominence (CP).

Caractéristique	Formule
Uniformité	$\sum_{i=0}^{NM-1} P_i^2(z_i)$
Cluster Prominence	$\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i + j - ux - uy)^4 P(i, j)$

Tableau 4.6 : Tableau des formules de caractéristiques rajoutées

Le tableau 4.7 illustre un exemple de deux vecteurs de caractéristiques qui combine quatorze caractéristiques des images n° 90 et n° 81.

Type	Descripteur	Masse maligne (image n° 90)	Masse bénigne (image n° 81)
Caractéristiques du 1er ordre	Moyenne	204.20886533365453	218.19402985074626
	Variance	431.6290289667555	522.379176018651
	Asymétrie	6.024787567978889	1.978878428742931
	Aplatissement	0.6318554853254963	22.61194029851099
	Entropie	0.5645143119098395	0.591390063138885
	Uniformité	0.7091954084177982	0.5981840482748391
	Régularité	0.954077216657895	0.9580811492705441
Caractéristiques Haralick	Moment angulaire du second degré	410638.0306503272	6597425.397156037
	Contraste	98.66046495086519	67.64965005312887
	Corrélation	6.791509738296413E-14	6.921781153568082E-16
	Variance	1.4719135199068676E13	1.4446249435556775E15

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

	Moment inverse	0.9877244496634364	0.982315581402438
	Entropie	-4376.407084450989	-23785.012312723098
	CP	1.1044719465718243E22	4.7658835523760595E24

Tableau 4.7 : Deux exemples de vecteurs de caractéristiques

4.3.2.2 Classification

Après le calcul des caractéristiques à partir les 56 régions d'intérêts, les vecteurs de caractéristiques sont envoyés aux six classificateurs différents : BN : *Bayes Network*, MLP : *Multi Layer Perceptron*, DT : *Decision Table*, DS : *Decision Stump*, J48 et RF : *Random Forest*, pour décider à la fin si la masse est maligne ou bénigne.

4.3.2.3 Résultats et discussion

Cinq métriques d'évaluations sont utilisées afin d'évaluer notre approche (MC, précision, sensibilité, spécificité et AUC), les résultats obtenus sont présentés dans le tableau 4.8.

Tous les valeurs de AUC sont supérieures à 0.5 ce qui confirme que la classification est bonne.

La capacité de déterminer correctement les cas bénins est bonne, elle est entre 74.3% et 80% ; et la capacité de déterminer correctement les cas malins est dans l'intervalle [59% - 91.6%], le taux de précision minimale est presque égale 70% et le maximale égale 80.4%.

	MC		Précision	Sensibilité	Spécificité	AUC
BN	7	14	0.75	0.666	0.8	0.745
	7	28				
MLP	8	13	0.696	0.590	0.764	0.707
	9	26				
DT	9	12	0.75	0.705	0.769	0.7
	5	30				
DS	10	11	0.804	0.804	0.772	0.652
	1	34				
J48	10	11	0.804	0.916	0.772	0.697
	1	34				
RF	10	11	0.714	0.647	0.743	0.695

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

	6	29				
--	---	----	--	--	--	--

Tableau 4.8 : Métriques d'évaluation utilisées

Les résultats obtenus ainsi que ceux de Jehlol et al. (Jehlol et al., 2015) sont affichés dans le tableau 4.9 et la figure 4.6.

La comparaison entre les deux travaux est fondée sur plusieurs critères, comme l'utilisation de la même base d'images MIAS (21 malignes et 35 bénignes), les caractéristiques statistiques et les mêmes classifieurs (DT, J48, DS, RF, MLP, BN).

Nous avons présenté une comparaison sur la performance des deux types de caractéristiques ; Jehlol et al. (Jehlol et al., 2015) ont obtenu un meilleur résultat (66.07%) lors ce qu'ils ont utilisé le classifieur « RF ».

Nous avons obtenus des meilleurs taux de précision en utilisant le classifieur « Decision Stump» ou bien le « J48 » avec 80.36%.

Classifieur	Caractéristiques	% de la précision
BN	Méthode de Jehlol et al. (Jehlol et al., 2015)	55.36 %
	Méthode proposée	75 %
MLP	Méthode de Jehlol et al. (Jehlol et al., 2015)	58.93 %
	Méthode proposée	69.64 %
DT	Méthode de Jehlol et al. (Jehlol et al., 2015)	55.36 %
	Méthode proposée	75 %
DS	Méthode de Jehlol et al. (Jehlol et al., 2015)	50%
	Méthode proposée	80.36%
J48	Méthode de Jehlol et al. (Jehlol et al., 2015)	48.21%
	Méthode proposée	80.36%
RF	Méthode de Jehlol et al. (Jehlol et al., 2015)	66.07%
	Méthode proposée	71.43%

Tableau 4.9 : Le taux de précision de la méthode proposée et la méthode de Jehlol et al. (Jehlol et al., 2015).

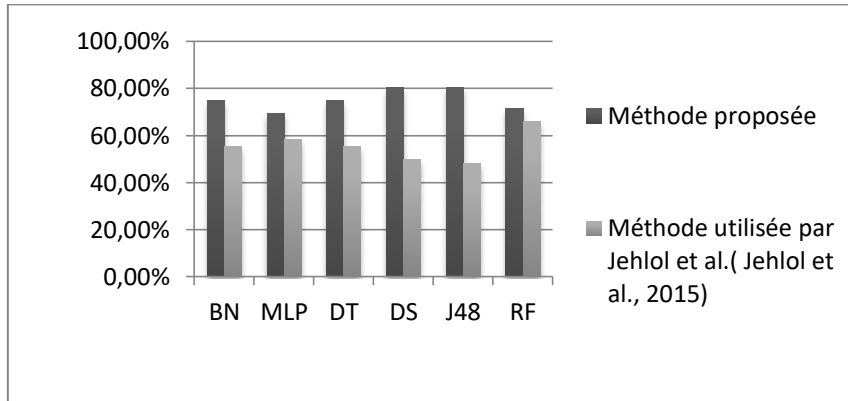


Figure 4.6 : Histogramme du taux de classification de la méthode proposée et la méthode de Jehloul et al. (Jehloul et al., 2015).

Le tableau 4.10 résume les quatorze caractéristiques du premier et du deuxième ordre utilisées dans notre approche et les cinq caractéristiques du second ordre uniquement, qui sont utilisées par Jehloul et al. (Jehloul et al., 2015) .

Caractéristiques	Méthode proposée	Méthode de Jehloul et al. (Jehloul et al., 2015)
Moyenne	x	
Variance	x	
Asymétrie	x	
Aplatissement	x	
Entropie	x	
Uniformité	x	
Régularité	x	
Moment angulaire du second degré	x	x
Contraste	x	x
Corrélation	x	x
Variance	x	x
Moment des différences inverses	x	
Entropie	x	x

CP	x	
----	---	--

Tableau 4.10 : Les caractéristiques utilisées

D'après les résultats obtenus, nous constatons que le taux de précision de notre approche est toujours meilleur par rapport à l'approche de Jehlol et al. (Jehlol et al., 2015) quelque soit le classifieur utilisé, donc notre combinaison de caractéristiques du premier et deuxième ordre renforce le taux de classification correcte des masses mammaires et donne de bons résultats malgré le que nombre d'image utilisé n'est pas volumineux.

Nous affirmons que le fusionnement de deux types de caractéristiques statistiques améliore la capacité de différencier correctement les cas de tumeurs malignes et bénignes.

4.3.3 Classification des masses mammaires basée sur les attributs statistiques du premier ordre

Dans les travaux précédents, nous avons montré que la combinaison des caractéristiques statistiques du premier et du deuxième ordre augmente le taux de la classification correcte quel que soit la technique de classification utilisée.

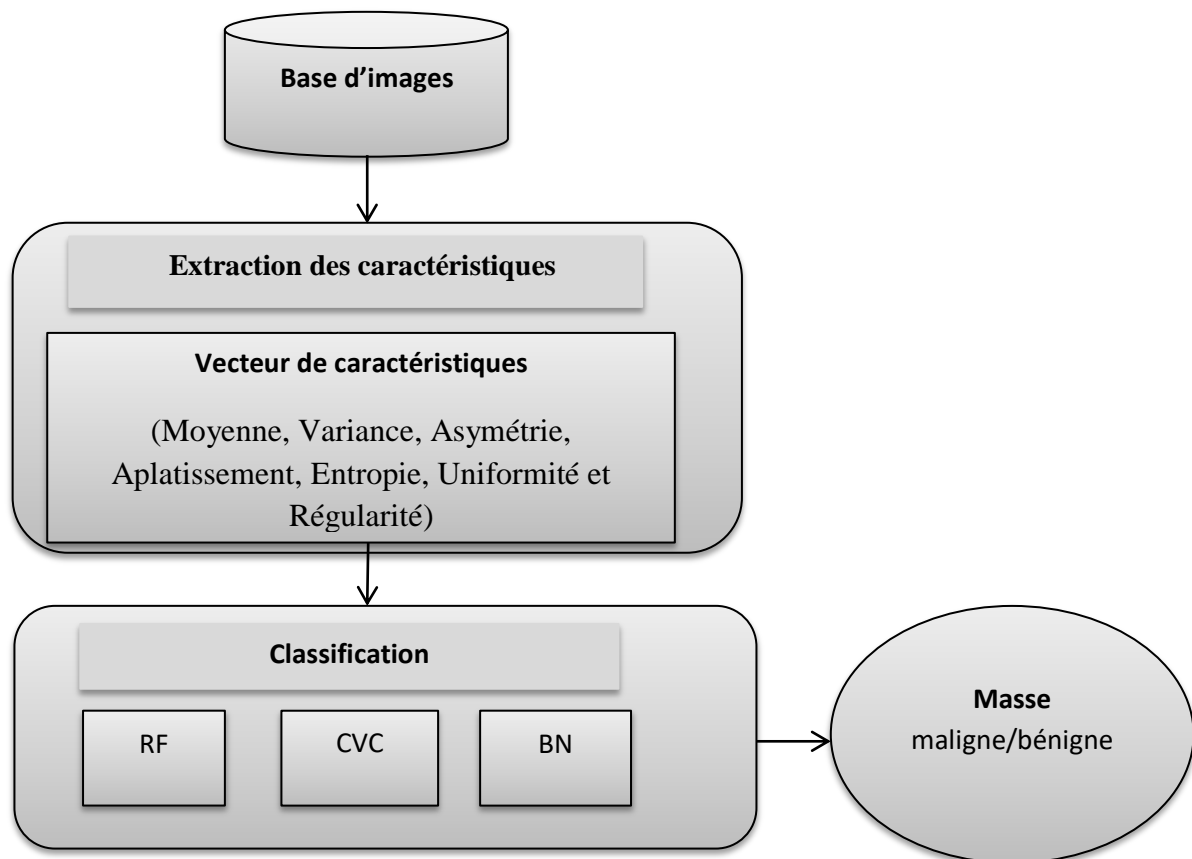


Figure 4.7 : Schéma de la méthode proposée

Dans ce travail qui a fait l'objet de la production scientifique de (*Boudraa et al., 2017c*), notre objectif reste toujours le même ç-à-d. classer les masses mammaires (maligne/bénigne), et là nous nous basons uniquement sur les caractéristiques du premier ordre pour confirmer notre hypothèse qui dit que ce type de caractéristique est vraiment l'influenceur sur le taux de classification correcte par rapport les descripteurs de Haralick. La figure 4.7 illustre le schéma de notre méthode proposée.

4.3.3.1 Extraction des caractéristiques

Dans la partie «Extraction des caractéristique», nous nous somme focalisés sur les mesures de texture du premier ordre, rappelons qu'elles se basent sur l'histogramme des niveaux de gris de l'image qui représente la densité de probabilité. Les principaux avantages de cette approche résident dans sa simplicité (Alsheh Ali, 2015) et son efficacité, c'est pourquoi ces caractéristiques sont très utilisées en imagerie médicale.

Nous avons calculé sept caractéristiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité et la régularité) à partir de la région d'intérêt. Le tableau 4.11 montre un exemple des valeurs de deux descripteurs de caractéristiques des images n° 92 et n° 15.

Type	Descripteur	Masse maligne (image n° 92)	Masse bénigne (image n° 15)
Caractéristiques du premier ordre	Moyenne	169.41006842619745	190.00155424308363
	Variance	571.7675285504556	518.6921498769112
	Asymétrie	1.6138058173482523	2.026946876407271
	Aplatissement	17.680348762949155	-22.901807168766233
	Entropie	0.635149755630867	0.5893466598198727
	Uniformité	0.5616699048467787	0.6011666724021036
	Régularité	0.9598581562783026	0.9579386823956294

Tableau 4.11 : Exemple de deux vecteurs de caractéristiques des images n° 92 et n° 15

4.3.3.2 Classification

Dans l'étape de classification, nous avons envoyé les vecteurs de caractéristiques au plusieurs type de classifieurs, mais nous avons gardé que les trois premiers qui ont donné les meilleurs résultats (RF : *Random forest*, CVC : *Classification-via-Clustering* et BN : *Bayes Network*).

4.3.3.3 Résultats et discussion

Le tableau 4.12 présente les résultats obtenus de la méthode proposée en fonction de sensibilité, spécificité et AUC.

Nous remarquons que le « AUC » dépasse toujours le 0.6, ce qui confirme que la classification est bonne. Le meilleur taux de sensibilité égal 68.1% avec une spécificité de 82.3%

	Sensibilité	Spécificité	AUC
RF	0.681	0.823	0.798
CVC	0.6	0.707	0.629
BN	0.666	0.843	0.772

Tableau 4.12 : Les critères d'évaluation de la troisième méthode

Les résultats de classification et les résultats des matrices de confusion de notre approche sont présentés dans (tableau 4.13 et tableau 4.14) ainsi que les résultats obtenus par la méthode de Jehlol et al. (Jehlol et al., 2015).

En se basant sur les deux tableaux précédents, nous observons que nos résultats sont meilleurs par rapport à l'autre méthode en utilisant les classifieurs BN et RF avec un taux de classification qui égale à avoisinant les 76.79%. Rappelons que Jehlol et al. (Jehlol et al., 2015) ont basé sur les cinq caractéristiques du premier ordre (moyenne, variance, asymétrie, aplatissement, entropie) ; l'apport de notre approche par rapport à l'autre travail est l'ajout des deux caractéristiques (uniformité, et régularité) à la combinaison des cinq caractéristiques.

Classifieur	Méthode	% de la précision	Le temps (seconde)
BN	Méthode de Jehlol et al. (Jehlol et al., 2015)	75 %	0.02s
	Méthode proposée	76.79 %	0.03 s
RF	Méthode de Jehlol et al. (Jehlol et al., 2015)	71.43 %	0.06 s

	Méthode proposée	76.79 %	0.05 s
CVC	Méthode de Jehlol et al. (Jehlol et al., 2015)	67.86 %	0.01 s
	Méthode proposée	67.86 %	0.05 s

Tableau 4.13 : Résultats de classification

Classifieur	Classe actuelle	Méthode proposée		Méthode de Jehlol et al. (Jehlol et al., 2015)	
		Maligne	Benigne	Maligne	Benigne
BN	Maligne	16	5	14	7
	Benigne	8	27	7	28
RF	Maligne	15	6	15	6
	Benigne	7	28	10	25
CVC	Maligne	9	12	11	10
	Benigne	6	29	8	27

Tableau 4.14 : Les résultats des matrices de confusion

Les résultats obtenus ont confirmé l'importance de la combinaison proposée qui a donné de bons résultats, donc le rajout des deux caractéristiques (Uniformité, et Régularité) a un apport positif sur le taux de la classification correcte.

4.4 Extraction de l'information de texture par les caractéristiques de Haralick et les motifs binaires locaux

Dans la plupart des systèmes d'aide au diagnostic, l'étape d'extraction des caractéristiques est basée sur (Oliveira et al., 2015) :

- La géométrie qui décrit la forme de la région de la lésion, par exemple, la rondeur de cette région et,

- La texture qui décrit l'aspect de la région de la lésion, par exemple, la répartition des niveaux de gris de la région.

Comme les caractéristiques de texture aident à discriminer et à identifier les objets (Caroline and Vaijyanthi, 2016), les caractéristiques de texture sont l'orientation de ce travail.

4.4.1 Représentation profonde de la texture via l'opérateur LBP

Dans ce travail qui a fait l'objet de la production scientifique de (Boudraa et al., 2018), deux niveaux sont appliqués à la région d'intérêt pour avoir une représentation profonde de la texture. La transformation du motif binaire local (LBP) définit le premier niveau, et le fonctionnement de la matrice de cooccurrence de niveau de gris (GLCM) décrit le deuxième niveau. La figure 4.8 décrit le schéma général de notre méthode.

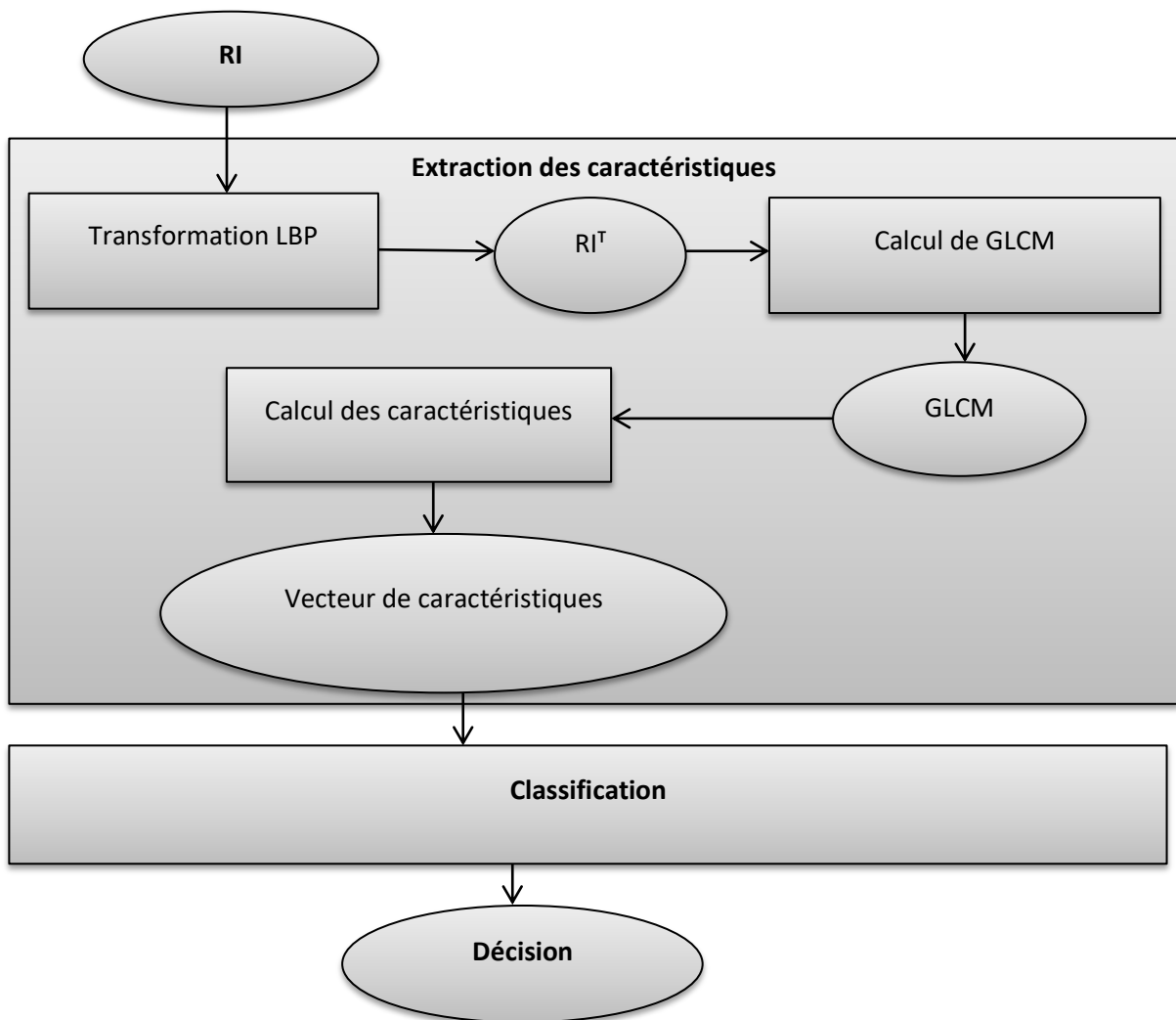


Figure 4.8 : Les étapes de la méthodologie proposée

4.4.1.1 La transformation via LBP

Les motifs binaires locaux ont été considérés comme une mesure de motifs invariants aux changements de niveaux de gris, qui offre une information quant à la texture d'une image ; LBP a garantie de bonnes performances en fonction de rapidité et de discrimination dans plusieurs travaux (Alsheh Ali, 2015).

Dans notre travail, nous appliquons la méthode LBP classique sur toutes les RIs acquises, nous obtenons donc des régions d'intérêts transformées (RI^Ts), qui sont prêtes à être envoyées à la deuxième étape « extraction des caractéristiques » ; la figure 4.9 montre un exemple de cette transformation.

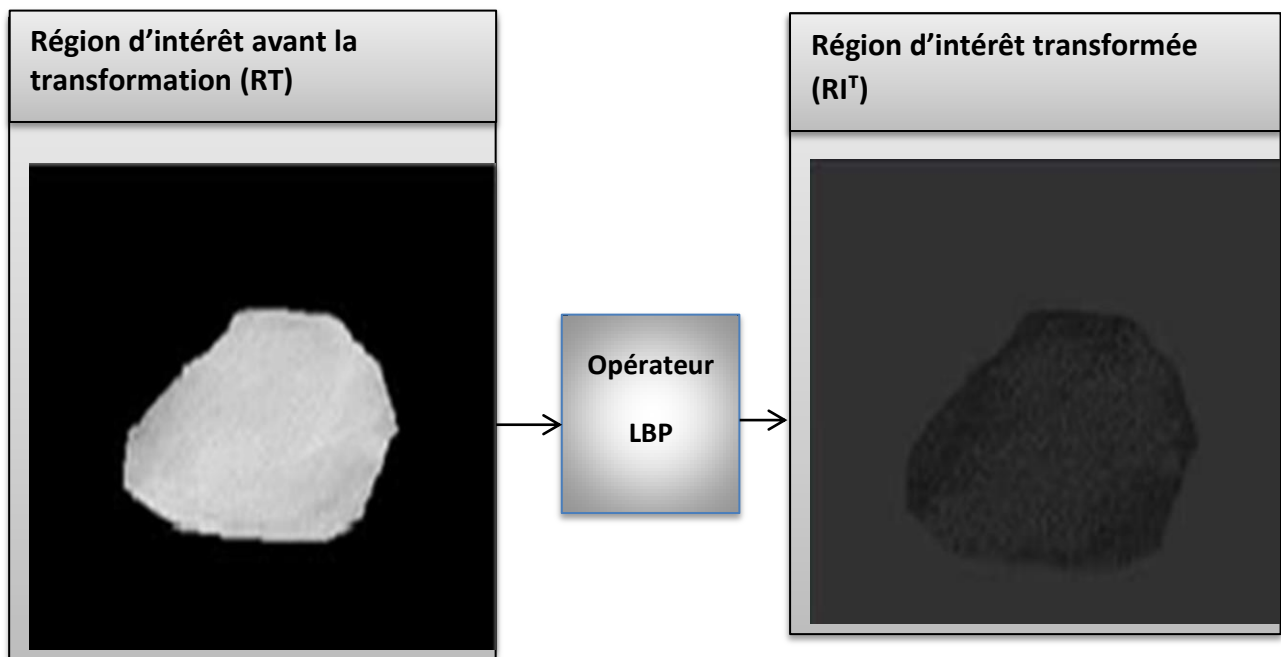


Figure 4.9 : Exemple d'application de l'opérateur LBP sur une RI

4.4.1.2 Extraction des caractéristiques

Pour mieux faire ressortir l'information de texture, nous calculons les matrices de co-occurrences (GLCM) à partir des RI^Ts, l'étape suivante sert à construire le vecteur de caractéristiques qui contient les caractéristiques du 2^{ème} ordre (le moment angulaire du second degré, le contraste, la corrélation, la variance, le moment des différences inverses et l'entropie) qui sont calculées à partir de la matrice GLCM. Le tableau 4.15 montre un exemple de deux vecteurs de caractéristiques des images n° 28 et n° 107.

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

Type	descripteur	Masse maligne (image n° 28)	Masse bénigne (image n° 107)
Descripteur de Haralick	Moment angulaire du second degré	59908.53693709501	577014.0666937797
	Contraste	81.0639306181852	51.68057534907986
	Corrélation	1.591618504934665E-12	3.676728794203147E-14
	Variance	6.276626842605773E11	2.719008235836071E13
	Moment des différences inverses	0.9901902953815024	0.9934685133169614
	Entropie	-1347.7768509862742	-5488.829032300574

Tableau 4.15 : Exemple de deux vecteurs de caractéristiques des images n°28 et n°107

4.4.1.3 Classification

Après la construction des vecteurs de caractéristiques, ils sont envoyés aux différents classifieurs (CVC : *Classifier Via Clustering*, BN : *Bayes Network*, DT : *Decision Table*, DS : *Decision Stump*, J48, et RT : *Random Tree*) afin de vérifier la validité de la méthode proposée.

4.4.1.4 Résultats et discussion

Afin d'évaluer notre travail, nous avons basé sur plusieurs métriques (MC, précision, sensibilité, spécificité et AUC) (voir tableau 4.16). Nous constatons que les taux de précision et de spécificité sont acceptables, par contre le taux de sensibilité est faible ; même les valeurs des AUC ne sont pas proche de la valeur 1 ce qui confirme que la classification n'est pas assez bonne, et ça est due premièrement au nombre faible des images, et deuxièmement à cause du manque de diversité de caractéristiques utilisées.

Classifieur	MC		Précision	Sensibilité	Spécificité	AUC
CVC	17	3	0.618	0.428	0.645	0.529
	4	31				
BN	21	0	0.625	0	0.625	0.444
	0	35				
DT	21	0	0.625	0	0.625	0.461
	0	35				
DS	21	0	0.571	0	0.603	0.378

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

	3	32				
J48	21	0	0.607	0	0.618	0.421
	1	34				
RT	12	9	0.643	0.529	0.692	0.6
	8	27				

Tableau 4.16 : Métriques d'évaluation.

Le tableau 4.17 et la figure 4.10 représentent les résultats obtenus par notre méthode par rapport à la méthode classique qui se base sur les régions d'intérêts originales (sans transformation) et qui utilise les mêmes caractéristiques calculées dans notre méthode proposée (le moment angulaire du second degré, le contraste, la corrélation, la variance, le moment inverse et l'entropie).

Classifieur	% de la precision	
	Méthode classique	Méthode proposée
CVC	55.36 %	60.71 %
BN	55.36 %	62.5 %
DT	55.36 %	62.5 %
DS	50%	57.14%
J48	48.22%	60.71%
RT	62.5%	64.29

Tableau 4.17 : Résultats de classification de la méthode proposée et la méthode classique.

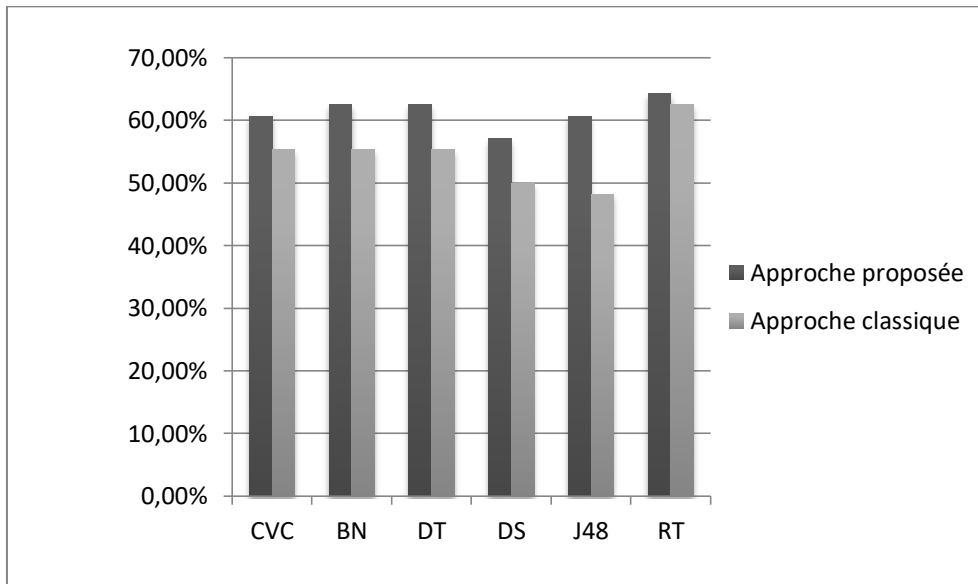


Figure 4.10 : Histogramme du taux de classification

Nous constatons que la méthode proposée donne de bons résultats par rapport à la méthode classique, ce qui prouve que l'opérateur LBP influence positivement et augmente le taux de classification correcte. Nous pouvons confirmer alors que le descripteur proposé a réellement apporté une bonne représentation de la texture.

4.5 Conclusion

Tout au long de ce chapitre, nous avons présenté une série d'expérimentations, qui ont pour objectif principal la classification des masses mammaires :

- Dans la première expérimentation, nous avons fusionné les descripteurs statistiques du premier et du deuxième ordre, et les résultats obtenus ont confirmés que cette combinaison a influencé positivement sur le taux de classification.
- Dans la deuxième et la troisième expérimentation, nous nous sommes adressés au même problème que le premier travail, c.-à-d. en se basant sur les caractéristiques qui sont purement statistiques, mais en ajoutant de nouvelles caractéristiques plus significatives.
- Dans la quatrième expérimentation, nous avons caractérisé la masse mammaire à l'aide d'une représentation profonde de texture via l'opérateur LBP et les descripteurs de Haralick. Les résultats obtenus sont prometteurs.

A noter que dans la phase d'évaluation nous avons utilisé des images tirées de la base de données MIAS à travers différentes techniques de classification, et nous avons basé sur plusieurs métriques d'évaluation.

Chapitre 04 : Caractérisation des masses mammaires à l'aide des attributs de texture dans les images normales

Des résultats satisfaisants ont été obtenus, ce qui indique que l'utilisation de l'information textuelle peut vraiment décrire les masses mammaires, ce qui garantit un bon taux de classification.

Chapitre 05

Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

5.1 Introduction

Dans le présent chapitre, nous continuons à travailler dans le même axe que la précédente contribution qui a comme objectif principal la classification des masses mammaires (bénigne/maligne).

L'idée de cette nouvelle contribution est de tester l'intégration de la technique de la super-résolution et les descripteurs de texture du premier et du troisième ordre dans le système d'aide au diagnostic dédié à la classification des masses mammaires dans les images mammographiques.

Afin de mettre en évidence la pertinence et l'efficacité de l'approche proposée, nous allons faire une étude comparative avec d'autres approches récentes rapportées dans la littérature.

5.2 La discrimination des masses mammaires par la super-résolution et les descripteurs de texture

La majorité des images médicales, acquises et affichées en niveaux de gris, sont très texturées, ce qui fait que l'examen des images médicales nécessite souvent une analyse et une interprétation de l'apparence des tissus. Il représente les variations locales d'intensité basées sur différentes propriétés de la texture : douceur, grossièreté, régularité et homogénéité. Comme la texture acquiert une telle importance, elle devient l'une des caractéristiques les plus puissantes utilisées dans l'analyse et la classification des images médicales (Hamoud et al., 2015). C'est pourquoi, dans cette étude, nous proposons une méthode qui est basée sur l'extraction des

caractéristiques de la texture et sur l'utilisation d'une technique qui améliore la discrimination des masses mammaires, puis nous appliquerons différentes techniques de classification afin de déterminer la classe de la masse (maligne/bénigne). Les différentes étapes de l'approche proposée seront présentées dans cette section, et elles sont résumées dans la figure 5.1. Ce travail a fait l'objet de la production scientifique de (Boudraa et al., 2020).

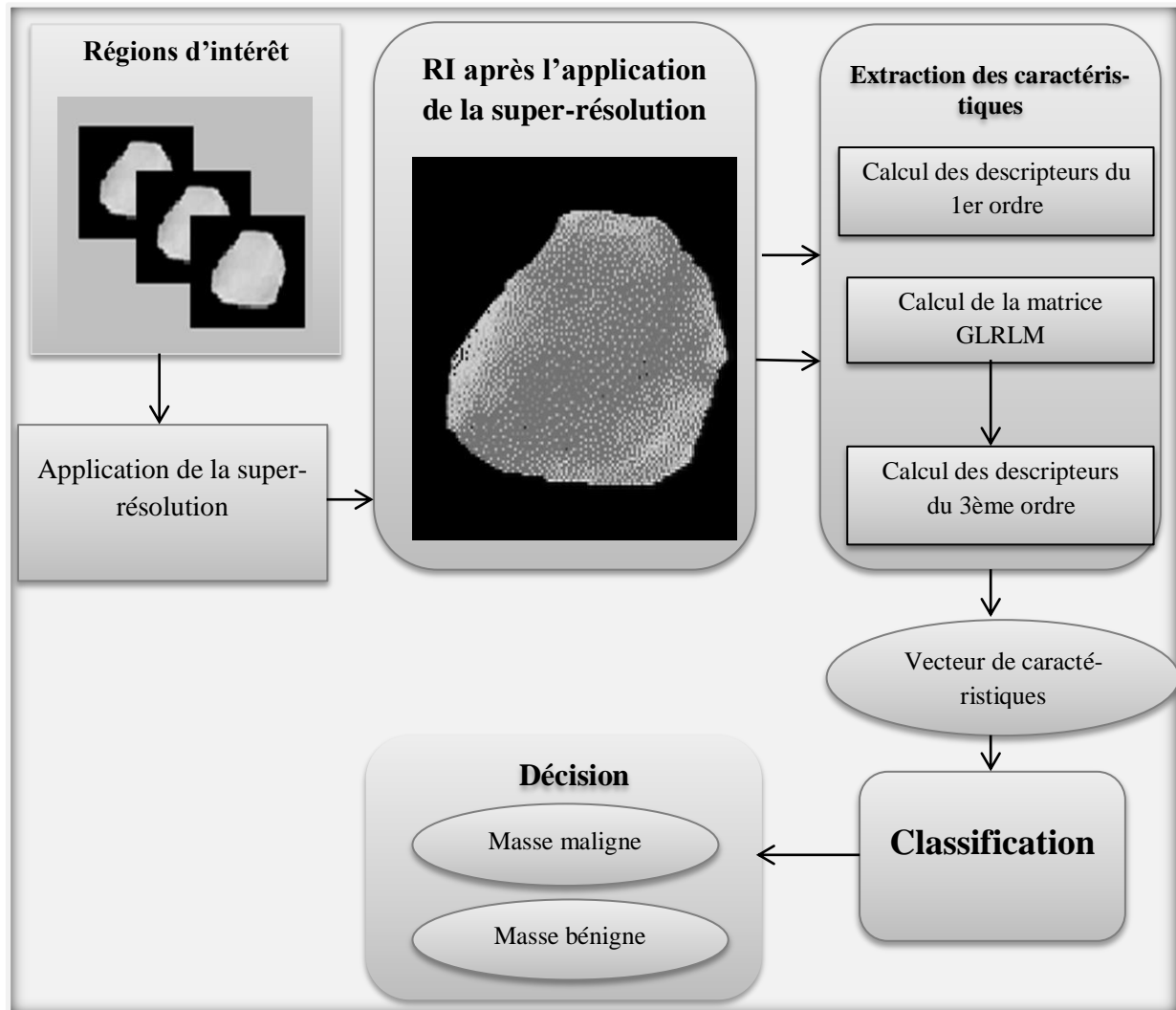


Figure 5.1 : Schéma général de la méthode proposée

5.2.1 Base d'images

Nous avons utilisé 93 images tirées de la base de données MIAS (*Mammographic Image Analysis Society*), où 54 sont de type bénin et 39 de type malin.

Après l'acquisition des images, nous avons les segmenté manuellement pour extraire les régions d'intérêts en se basant sur la position du centre de la masse (qui est déjà marqué au niveau des images dans la base MIAS) via l'outil ImageJ.

5.2.2 Application de la super-résolution

Dans ce travail, une nouvelle étape pour améliorer la discrimination des masses en utilisant la reconstruction à super-résolution est introduite. Rappelons que la super-résolution (SR) est une technique qui permet d'obtenir une ou plusieurs images à haute résolution (HR) à partir d'une ou plusieurs observations à basse résolution (BR).

La SR est un algorithme qui vise à fournir des détails plus fins que la grille d'échantillonnage d'un dispositif d'imagerie donné en augmentant le nombre de pixels par unité de surface dans une image (Nasrollahi and Moeslund, 2014).

Afin d'intégrer la technique de la super-résolution dans notre système, nous avons utilisé la méthode de super-résolution de Chung and Nagy (Chung and Nagy, 2008) qui est disponible sous la forme d'un plugin via l'outil ImageJ qui s'appelle « *Parallel Super-Resolution* ».

Après l'acquisition des régions d'intérêts à basse résolution, nous faisons appel à la technique de la super-résolution à travers le plugin « *Parallel Super-Resolution* », nous utilisons ses paramètres par défaut, à la fin nous obtenons une image à haute résolution qui est prête à être utilisée dans la prochaine étape qui est l'extraction des caractéristiques. Les figures (5.2 et 5.3) montrent en détail comment la super-résolution est appliquée sur une région d'intérêt.

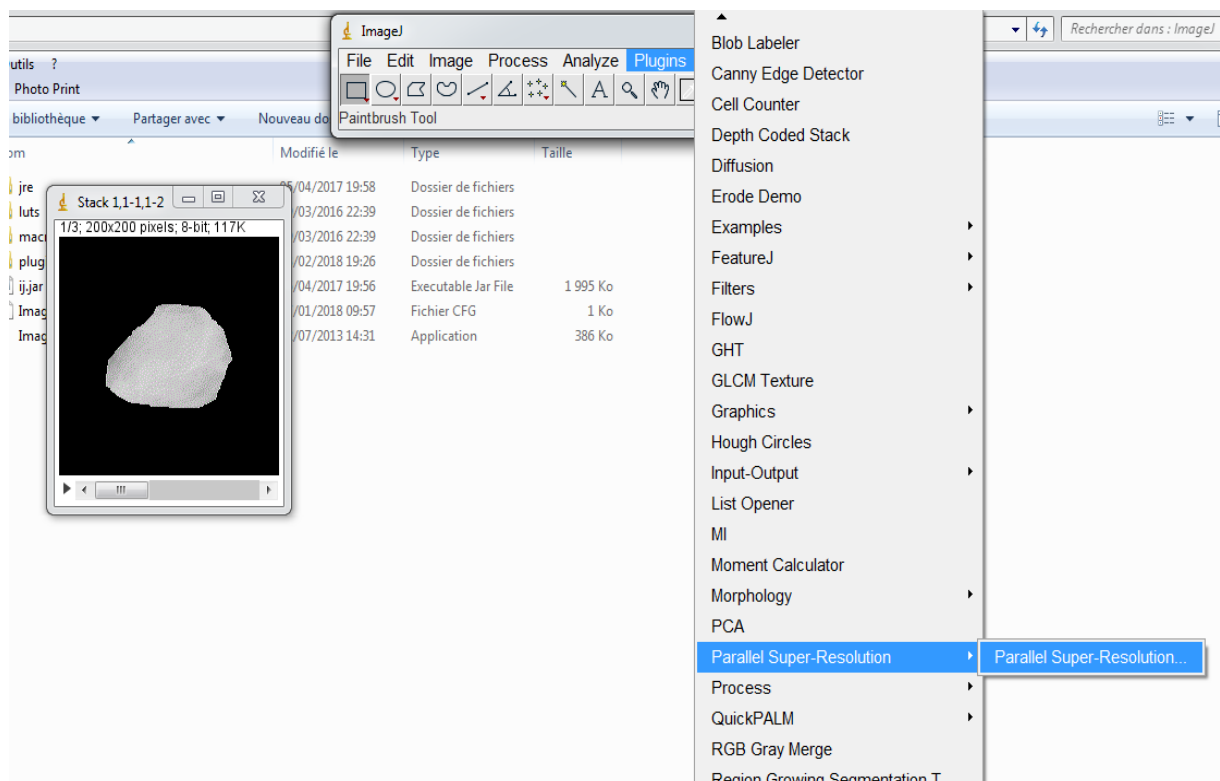


Figure 5.2 : Utilisation du plugin « Parallel Super-Resolution »

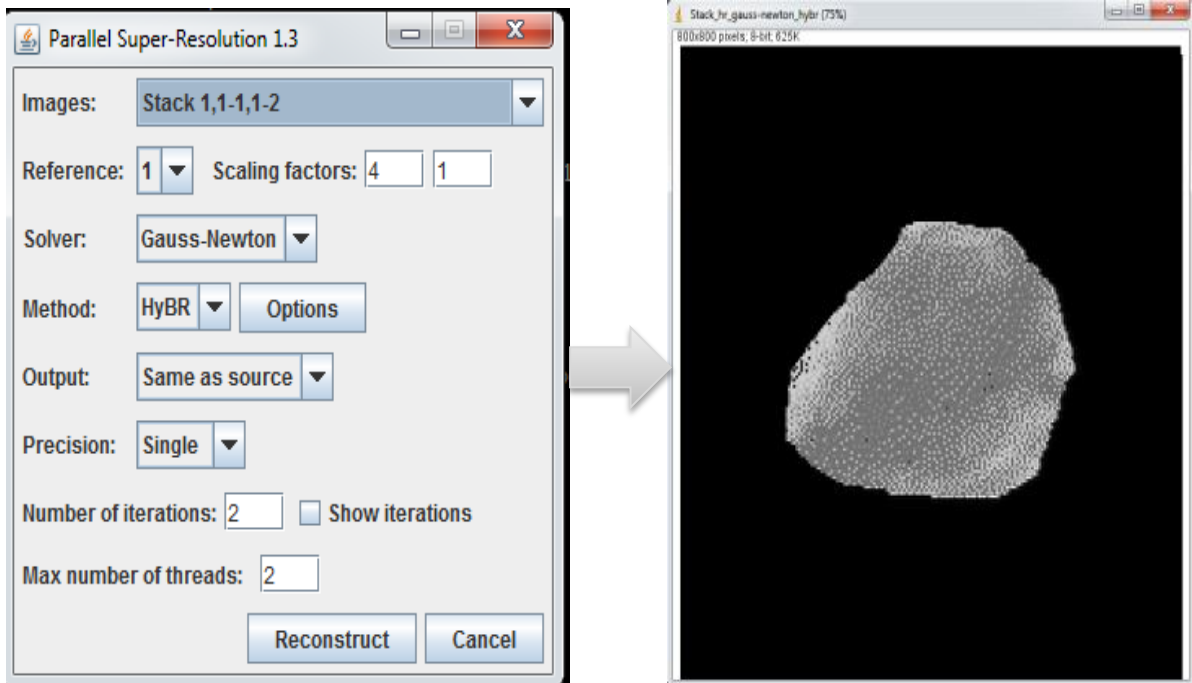
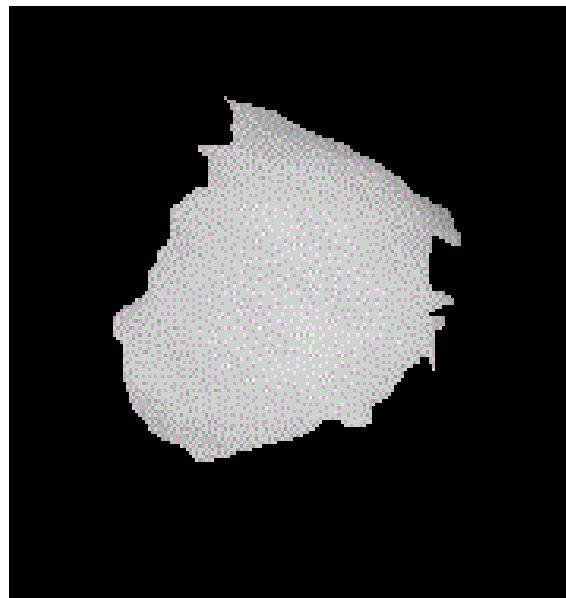


Figure 5.3 : Résultat de l'application de la SR

La figure 5.4 illustre un exemple de deux types de régions d'intérêt, où la première est bénigne et la deuxième est maligne, avant et après l'application de la méthode de la super-résolution.



(a)



(b)

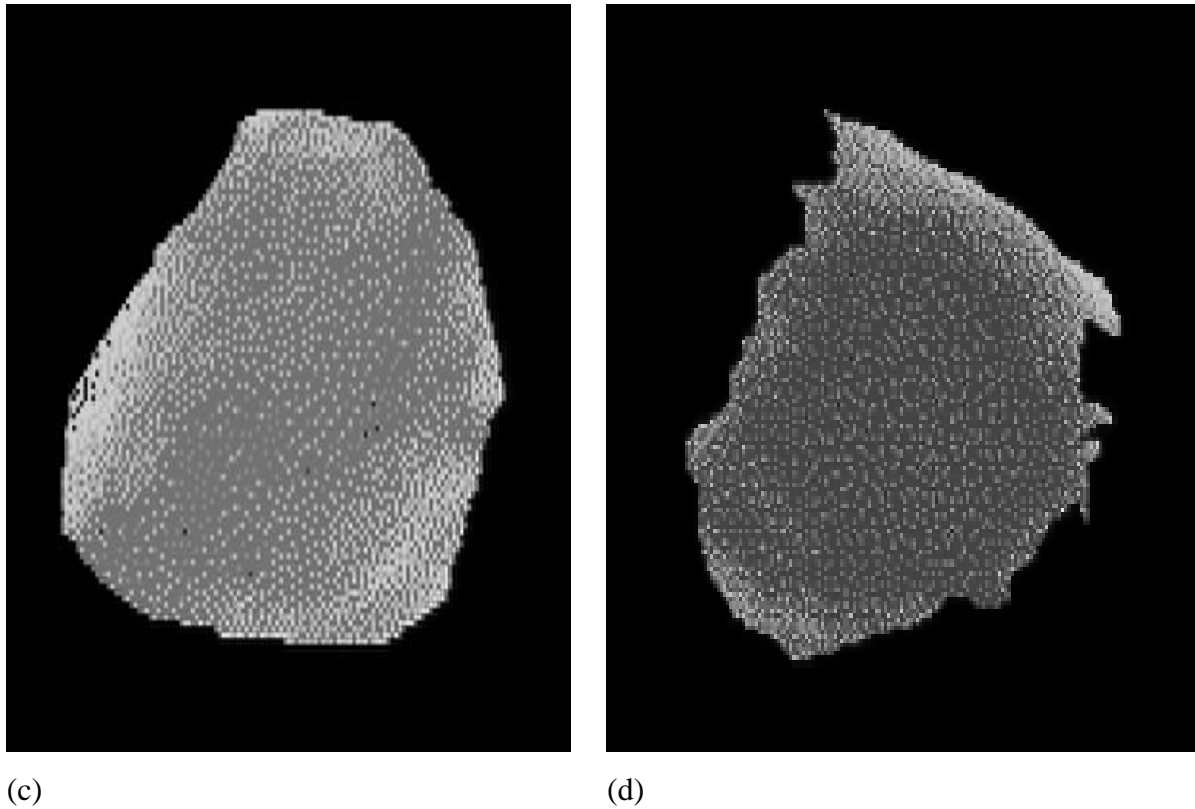


Figure 5.4 : Exemple d'application de la super-résolution : (a) masse bénigne avant SR (b) masse maligne avant SR (c) masse bénigne après SR (d) masse maligne après SR

5.2.3 Extraction des caractéristiques

Devant le grand nombre de descripteurs utilisés dans le domaine de classification des images mammaires, nous avons choisi de travailler sur les descripteurs de texture.

Comme nous avons déjà souligné dans le chapitre précédent, la texture du sein dans les images mammographiques joue un rôle très important dans la classification des masses (Shastri et al. 2018), c'est pourquoi les caractéristiques de texture sont largement utilisées dans de nombreuses approches de détection de masse des seins.

Notre étude se concentre sur l'évaluation de l'utilisation de deux méthodes statistiques de texture : les statistiques du premier et du troisième ordre ; rappelons que les statistiques du premier ordre estiment les caractéristiques qui reposent uniquement sur les valeurs des pixels individuels de l'image et ignorent l'interaction spatiale entre les pixels de l'image ; ce genre de descripteurs a montré son efficacité pour la distinction de type des masses dans nos travaux précédents.

Les descripteurs statistiques d'ordre supérieur, se basent sur la matrice de longueur de niveau de gris (GLRLM) qui est une méthode appliquée avec succès pour représenter la texture.

Nous calculons les matrices GLRLM à partir des régions d'intérêt, en définissant le nombre de niveaux de gris qui désigne la hauteur de la matrice GLRLM et la direction θ (0° , 45° , 90° ou 135°) ; Rappelons que la case (i, j) de la matrice désigne le nombre de segments de longueur i et de niveau d'intensité j dans une direction θ (Thibault, 2009) ; un exemple de construction du GLRLM est montré dans la figure 5.5, où la direction $\theta = 0^\circ$. GLRLM caractérise la grossièreté d'une texture, ce qui correspond d'avantage à la définition d'une texture homogène (Thibault, 2009). GLRLM paraît la plus adaptée pour l'utiliser dans notre système d'aide au diagnostic proposé.

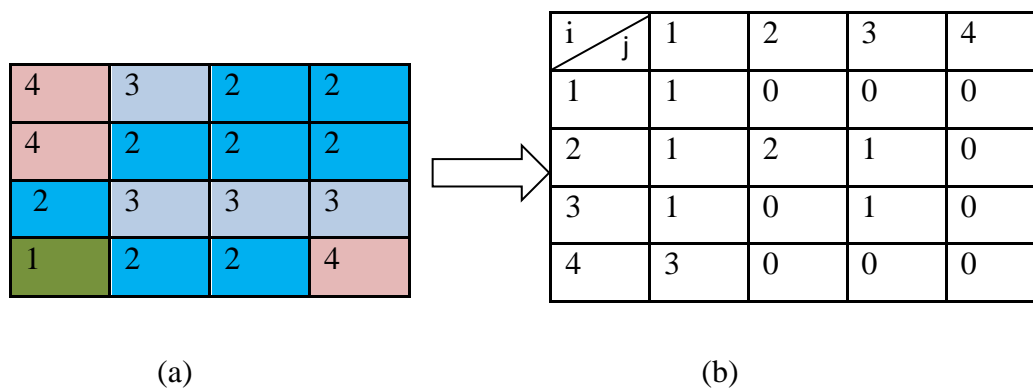


Figure 5.5 : Construction de la matrice de longueur de niveau de gris (a : matrice originale, b : matrice GLRLM)

Les caractéristiques statistiques du premier ordre sélectionnées à partir de l'image originale sont : la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité et la régularité.

Les caractéristiques statistiques d'ordre supérieur sélectionnées à partir de la matrice GLRLM sont : SRE : *Short Run Emphasis*, LRE : *Long Run Emphasis*, LGLRE : *Low gray level Run Emphasis*, HGLRE : *High Gray Level Run Emphasis*, SRLGLE : *Short Run Low Gray Level Emphasis*, SRHGLE : *Short Run High Gray level Emphasis*, LRLGLE : *Long Run Low Gray Level Emphasis*, LRHGLE : *Long Run High gray Level Emphasis*, GLN : *Gray Level No uniformity*, RLN : *Run length no uniformity*, RP : *Run Percentage*.

Les formules de ces caractéristiques ainsi que leurs définitions sont présentées dans le tableau 5.1.

Descripteur	Formule (Filipcuk et al., 2012)	Définition (Xu et al., 2004)
SRE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{j^2}$	Mesure la distribution des courts segments. Il est fortement dépendant de l'occurrence des courts segments et devrait être important pour les textures fines.

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

LRE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N p(i,j) j^2$	Mesure la distribution des longs segments. Le LRE est fortement dépendant de l'occurrence des longs segments et devrait être important pour les textures structurales grossières.
LGLRE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{i^2}$	Mesure la distribution des valeurs de niveaux de gris faibles. La valeur la plus grande indique que les valeurs de niveaux de gris sont faibles dans l'image.
HGLRE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N p(i,j) i^2$	Mesure la distribution des valeurs de niveaux de gris élevées. La valeur la plus grande indique que les valeurs de niveaux de gris sont élevées dans l'image.
SRLGLE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{i^2 j^2}$	Mesure la distribution conjointe des courts segments et des faibles valeurs de niveaux de gris. La valeur la plus grande indique de nombreux courts segments et des valeurs faibles de niveau de gris dans l'image.
SRHGLE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) i^2}{j^2}$	Mesure la distribution conjointe des courts segments et des valeurs élevées de niveaux de gris. La valeur la plus grande indique de nombreux courts segments et des valeurs élevées de niveau de gris dans l'image.
LRLGLE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) j^2}{i^2}$	Mesure la distribution conjointe des longs segments et des faibles valeurs de niveaux de gris. La valeur la plus grande indique de nombreux longs segments et des valeurs de niveau de gris plus faibles dans l'image.
LRHGLE	$\frac{1}{nr} \sum_{i=1}^M \sum_{j=1}^N p(i,j) i^2 j^2$	Mesure la distribution conjointe des longs segments et des valeurs élevées de niveaux de gris. La valeur la plus grande indique de nombreux longs segments et des valeurs élevées de niveau de gris dans l'image.
GLN	$\frac{1}{nr} \sum_{i=1}^M \left(\sum_{j=1}^N p(i,j) \right)^2$	Mesure la similarité des valeurs des niveaux de gris sur toute l'image. Le GLN est considéré comme faible si les valeurs des niveaux de gris sont identiques sur toute l'image.
RLN	$\frac{1}{nr} \sum_{j=1}^N \left(\sum_{i=1}^M p(i,j) \right)^2$	Mesure la similarité des longueurs de segments sur toute l'image. Le RLN est considéré comme faible si les longueurs des segments sont identiques sur toute l'image.
RP	$\frac{nr}{np}$	Mesure l'homogénéité et la distribution des segments d'une image dans une direction spécifique. La valeur la

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

		plus grande indique que la longueur des segments est 1 pour tous les niveaux de gris dans une direction spécifique.
--	--	---

Tableau 5.1 : Les caractéristiques d'ordre supérieur sélectionnées.

Le tableau 5.2 illustre un exemple de deux vecteurs de caractéristiques extraits à partir de deux régions d'intérêt l'une bénigne et la deuxième maligne (avec et sans l'application de la technique de la super-résolution).

		Descripteurs proposés	a) Masse bénigne		b) Masse maligne	
			Avant SR	Après SR	Avant SR	Après SR
Statistique du premier ordre	1	Moyenne	2.0059	1.3556	2.0028	0.8749
	2	Variance	3.6610	15.7285	3.5150	19.3390
	3	Aplatissement	6.6210	3.3315	6.9200	4.3836
	4	Asymétrie	- 13.8877	12.7418	- 16.8182	50.0128
	5	Entropie	0.4991	3.7542	0.4832	4.7561
	6	Uniformité	0.7431	0.0959	0.7523	0.0142
	7	Régularité	0.9503	0.9754	0.9493	0.9777
Statistiques d'ordre supérieur	8	SRE	0.6340	0.5862	0.5900	0.5986
	9	LRE	2.4175	4.5231	15.4354	26.3673
	10	GLN	7.5214	13.8336	6.7721	7.8938
	11	RP	0.1595	0.2492	0.0773	0.1664
	12	RLN	7.7367	202.9192	6.1087	314.6366
	13	LGLRE	0.0424	0.0190	0.0603	0.0342
	14	HGLRE	374.6323	207.3370	364.8763	96.5961
	15	SRLGLE	0.0034	0.1700	0.0364	0.7600
	16	LRLGLE	243.2817	123.5257	224.8109	59.9512
	17	SRHGLE	1.6803	4.3834	14.7148	26.3239
	18	LRHGLE	30912.3107	2373.4983	30199.4957	441.7784

Tableau 5.2 : Exemple de deux vecteurs de caractéristiques avec et sans l'utilisation de la super-résolution.

5.2.4 Phase d'apprentissage

La phase d'apprentissage, se décompose en deux principales parties :

1. Sans la super-résolution

Dans cette partie, nous nous basons sur les 93 images originales (sans aucune modification), et nous suivons les étapes suivantes :

- Création de la base d'apprentissage « base-1.arff ».
- Calcul des caractéristiques.
- Construction des vecteurs de caractéristiques.
- Remplissage de la base d'apprentissage (voir figure 5.6).

2. Avec la super-résolution

Dans cette partie, nous suivons les étapes suivantes :

- Application de la technique de la super-résolution sur les 93 images originales.
- Création de la base d'apprentissage « Base-SR.arff ».
- Calcul des caractéristiques sur la collection des images prétraitées par la super-résolution.
- Construction des vecteurs de caractéristiques.
- Remplissage de la base d'apprentissage (voir figure 5.7).

Le principe est le même dans les deux étapes, une fois le calcul de toutes les caractéristiques est obtenu, nous les fusionnons dans un seul vecteur qui contient les caractéristiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité, la régularité) et les caractéristiques d'ordre supérieur (SRE, LRE, LGLRE, HGLRE, SRLGLE, SRHGLE, LRLGLE, LRHGLE, GLN, RLN et RP). Chaque vecteur de caractéristiques construit sera rajouté dans la base d'apprentissage pour l'envoyer à la dernière étape « la classification ».

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

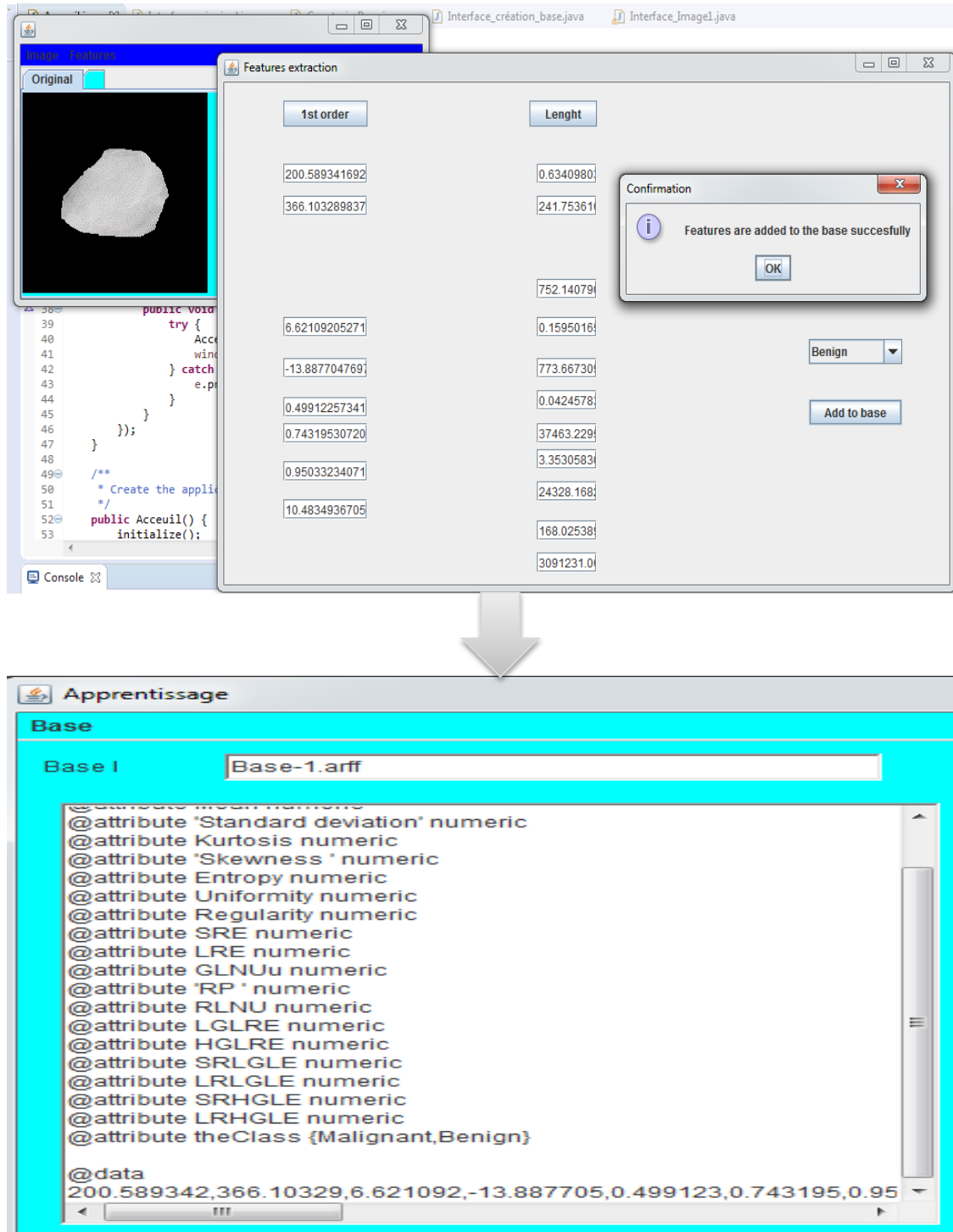


Figure 5.6 : Apprentissage sans l'utilisation de la SR

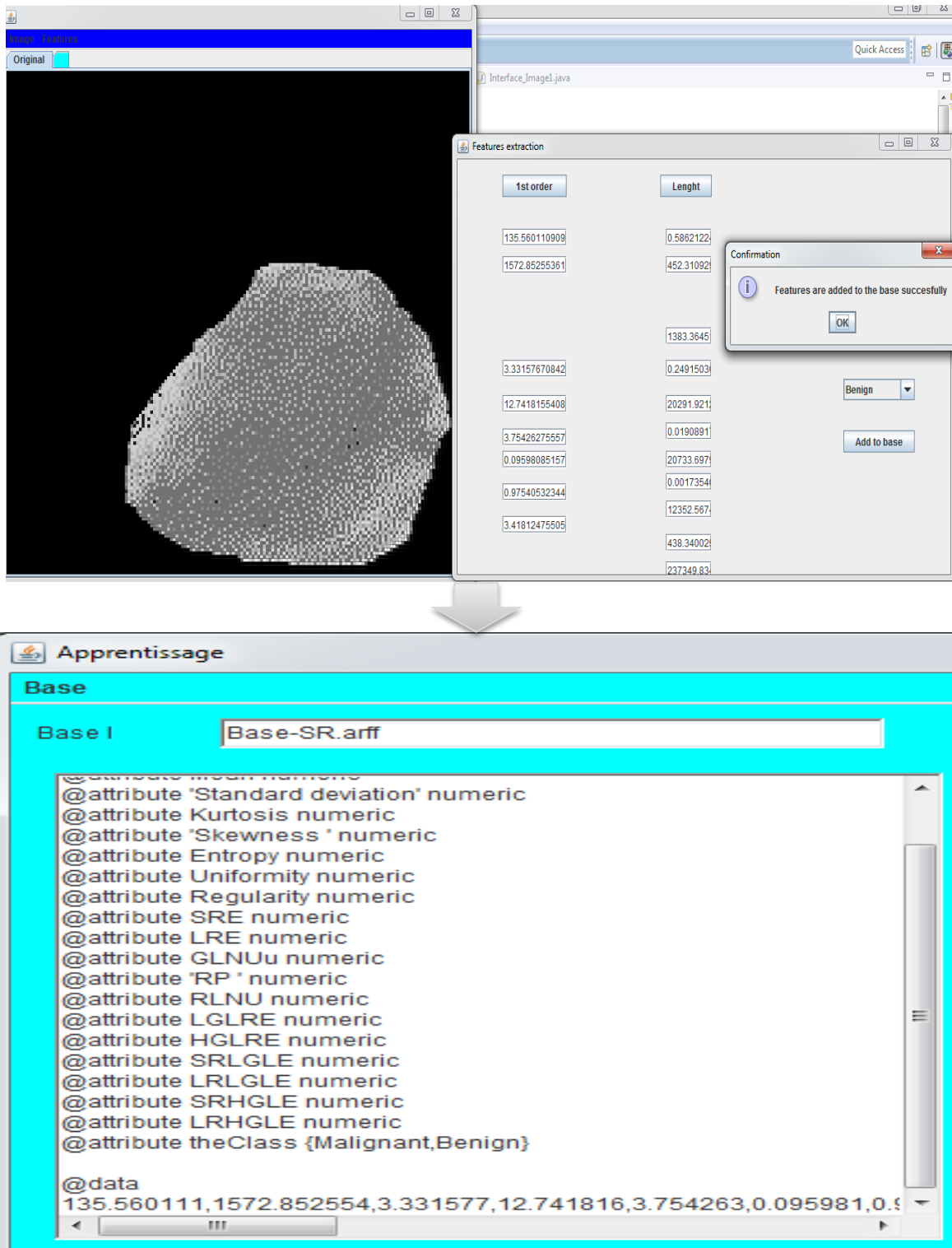


Figure 5.7 : Apprentissage avec l'utilisation de la SR

5.2.5 Classification

Afin de visualiser clairement les performances de l'approche suggérée, dans l'étape de « classification », nous avons testés plusieurs techniques de classifications, et nous avons gardé que les classifieurs qui ont donné les meilleurs résultats.

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

Nous avons utilisé neuf classifieurs disponibles via l'outil Weka (Simple Cart, Decision Table, Threshold Selector, Random Forest, Multi Class Classifier, Classification-Via-Regression, LWL, Simple Logistic et Logistic), où la valeur de la validation croisée k-fold =10.

5.2.6 Résultats et discussion

Afin de fournir une évaluation à notre approche, nous nous sommes basés sur trois métriques d'évaluation : 'la précision' qui est la capacité de différencier correctement les cas de tumeurs malignes et bénignes, 'la sensibilité' qui est la capacité de déterminer correctement les cas malins, et 'la spécificité' qui est la capacité de déterminer correctement les cas bénins.

Le tableau 5.3 présente les résultats de la classification « avec et sans l'utilisation de la méthode de super-résolution » en termes de précision, sensibilité et spécificité.

Classifieur	Précision sans SR	Précision avec SR	sensibilité sans SR	sensibilité avec SR	Spécificité sans SR	Spécificité avec SR
Simple cart	77.4%	82.7%	73.6%	81%	80%	83.9%
Decision table	77.4%	84.9%	80%	96.2%	76.1%	80.3%
Threshold selector	84.9%	89.2%	83.7%	89.1%	85.7%	89.2%
Rotation forest	89.2%	94.6%	85.3%	94.7%	92.3%	94.5%
Multi class classifier	80.6%	89.2%	78.3%	89.1%	82.1%	89.2%
Classification-Via-Regression	84.9%	94.6%	85.7%	97.2%	84.4%	92.9%
LWL	82.7%	92.4%	82.8%	94.4%	82.7%	91.2%
Simple logistic	87 %	96.7%	88.5%	100%	86.2%	94.7%
logistic	80.6%	89.2%	76.9%	89.1%	83.3%	89.2%

Tableau 5.3 : Récapitulatif des performances des différents classifieurs

Les résultats expérimentaux obtenus témoignent que la super-résolution améliore les taux de classification (précision, sensibilité et spécificité) quel que soit le classifieur utilisé, et d'après ces résultats, nous constatons que le taux de la précision le plus élevé (**96,7 %**) est atteint dans

Chapitre 05 : Amélioration de la discrimination des masses mammaires par l'intégration de la super-résolution et les descripteurs de texture

le cas où la technique de la super-résolution est appliquée en utilisant le classifieur 'Simple Logistic', avec **100 %** de sensibilité et **94,7 %** de spécificité.

Afin de valider la stratégie de classification proposée et afin de ressortir ses points forts, et ses points faibles, nous avons la comparé avec d'autres approches de l'état de l'art, le tableau 5.4 présente cette comparaison qui est basée sur la précision, où Görgel et al. (Görgel et al., 2015) ont appliqué la transformée en ondelettes stationnaire sur les régions d'intérêt, ensuite, ils ont extrait des descripteurs basés sur la forme et le niveau de gris ; Seryasat et Haddadnia (Seryasat et Haddadnia, 2017) se sont focalisés sur les descripteurs de texture et les descripteurs morphologiques ; Sheba et Raj (Sheba et Raj, 2018) ont utilisé les descripteurs calculés à partir de la matrice 'GLRLM', les descripteurs de forme, les descripteurs de GLCM et les descripteurs statistiques du premier ordre ; Braz et al. (Braz et al., 2018) ont diversifié l'analyse des régions de masse par les descripteurs de la géométrie géostatistique ; Jehlol et al. (Jehlol et al., 2015) ont évalué les descripteurs de texture du premier et du deuxième ordre, et ils ont obtenue des meilleurs résultats avec les descripteurs statistiques du second ordre.

Tous les travaux sélectionnés, dans cette comparaison, se sont concentrés sur la classification des régions d'intérêt dans les images mammographiques, extraites de la base de données MIAS.

Nous remarquons que le pourcentage de précision obtenu par notre approche est supérieur à celui de la plupart des approches mentionnées, cela signifie que l'intégration de la technique de super-résolution dans un système d'aide au diagnostic offre une puissance de discrimination entre les masses malignes et bénignes dans les images mammographiques.

Référence	Type de région	Base de données	Précision
Jehlol et al. (2015)	Maligne/Benigne/Normale	MIAS	85.8%
Görgel et al. (2015)	Maligne / Benigne	MIAS	90.1%
Seryasat and Haddadnia (2017)	Maligne / Benigne	MIAS	93%
Sheba and Raj (2018)	Maligne / Benigne /Normale	MIAS	96%
Braz Junior, et al. (2018)	Masse/Normale	MIAS	97.3%
Approche proposée	Maligne / Benigne	MIAS	96.7%

Tableau 5.4 : La précision de l'approche proposée par rapport à celle des autres travaux

Ainsi ; dans cette étude, nous pouvons affirmer que :

- L'ajout d'une nouvelle étape entre l'extraction de la région d'intérêt et l'étape de classification en utilisant une reconstruction à super-résolution améliore la qualité de la discrimination entre les masses mammaires.
- La combinaison des dix-huit caractéristiques statistiques de texture du premier ordre et d'ordre supérieur qui comprennent la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité, la régularité, SRE, LRE, LGLRE, HGLRE, SRLGLE, SRHGLE, LRLGLE, LRHGLE, GLN, RLN et RP, permet d'obtenir des performances élevées dans la classification des lésions mammaires.
- Les résultats atteints sont prometteurs et ils montrent que l'approche proposée est bonne et qu'elle est capable de distinguer les masses malignes des masses bénignes dans les images mammographiques, avec une précision élevée de **96,7%** par rapport à la plupart des méthodes existantes.

5.3 Conclusion

Les systèmes d'aide de diagnostic peuvent sauver des vies en détectant le cancer du sein à son stade précoce. L'une des étapes les plus difficiles dans un système d'aide au diagnostic dédié aux images mammographiques est la reconnaissance des formes qui englobe l'extraction des caractéristiques et la classification des masses. De nombreuses études ont montrés que le bon choix de caractéristiques améliore la précision du système d'aide au diagnostic.

Dans ce chapitre, une nouvelle contribution a été apportée dans le problème de classification des masses mammaires, où nous nous sommes focalisés sur l'intégration de la technique de la super-résolution et l'utilisation de deux familles de caractéristiques de textures qui sont purement statistiques.

A travers notre étude, nous avons détaillé l'approche proposée et nous avons obtenu des résultats intéressants comparativement à la plupart des travaux similaires.

A partir des expérimentations menées, nous avons pu prouver que notre système a atteint de très bonnes performances par l'intégration de la technique de la super-résolution.

De ce fait, nous pouvons dire que la super résolution est la méthode appropriée à utiliser dans l'étape de la discrimination des masses mammaires afin d'améliorer les performances de la phase de reconnaissance des formes, et la précision du système d'aide au diagnostic.

Conclusion générale et perspectives

1. Récapitulatif

Ce travail de recherche s'inscrit dans le contexte médical et il tourne autour de la classification des masses mammaires dans un système d'aide au diagnostic médical.

La mammographie est largement utilisée comme la modalité d'image la plus efficace pour le dépistage précoce de la maladie du cancer du sein. Pourtant, la structure anatomique dense du sein et la similitude entre les masses malignes et bénignes compliquent la tâche de la lecture des images mammographiques. C'est pourquoi les systèmes d'aide au diagnostic ont un rôle très important ; ils apportent une aide aux radiologues par un deuxième avis qui est capable de détecter des spécificités anatomiques des lésions dans les images mammographiques.

Les étapes primordiales dans un système d'aide au diagnostic sont l'extraction des caractéristiques et la classification. De nombreuses études ont montré que l'extraction automatique discriminante des caractéristiques améliore la précision de classification et garantit le bon fonctionnement d'un système d'aide au diagnostic. Le choix de caractéristiques efficaces est un problème qui nécessite des études pour obtenir une meilleure discrimination entre les masses malignes et bénignes. La texture est considérée comme le type de caractéristiques le plus utilisé pour les images médicales, c'est pourquoi, nous avons exploité ce type de caractéristique pour la discrimination entre les masses mammaires.

Dans le premier chapitre de ce mémoire, nous avons donné une vue globale sur les notions médicales concernant le cancer du sein, ainsi que la technique d'imagerie médicale « la mammographie ». Le second chapitre présente les différentes notions de base concernant les descripteurs de texture du premier, du deuxième et du troisième ordre. Il présente également les motifs binaires locaux comme descripteur de texture. Un autre volet dans ce chapitre est consacré à la présentation de la technique de la super-résolution dédiée à l'amélioration de la qualité des images. Dans le troisième chapitre, nous avons dressé un état de l'art des principaux descripteurs utilisés dans le cadre de la classification des régions d'intérêts dans les images mammographiques. La liste des descripteurs dressée a permis de mettre en évidence que les descripteurs choisis influence sur les performances de la classification, et

elle a confirmé que l'intégration des descripteurs de texture dans l'étape d'extraction des caractéristiques est très avantageuse. Parallèlement, nous avons également exposé quelques travaux qui sont basés sur la méthode de la super-résolution dans le domaine médical, et qui ont affirmé que cette méthode influence positivement sur la qualité des images médicale.

Lors du quatrième et cinquième chapitre, nous avons présenté nos contributions qui comportent plusieurs expérimentations, qui ont été testées sur des images mammographiques tirées à partir de la base MIAS (*Mammographic Image Analysis Society*).

2. Expérimentations réalisées :

Dans cette thèse, nous avons réalisé cinq expérimentations afin de guider le diagnostic du cancer du sein :

- **La première expérimentation « Hybridation des caractéristiques statistiques »** : consiste à fusionner les caractéristiques du premier et du deuxième ordre et les évaluer à travers quatre techniques de classification (DT, J48, RF et BN), le meilleur taux atteint de la classification correcte égal à **80.4%** en utilisant le classifieur « **J48** ».
- **La deuxième expérimentation « Hybridation des descripteurs du premier ordre et de Haralick »** : est une extension du première expérimentation; afin de bien caractériser les masses mammaires, nous rajoutons d'autres descripteurs statistiques dans l'étape d'extraction de caractéristiques, et pour valider la combinaison proposée, six classifieurs ont été utilisés (DT, J 48, DS, RF, MLP et BN), le meilleur taux de classification correcte atteint égal à **80.4%** en utilisant les classifieurs « **Decision Stump** » et « **J48** ».
- **La troisième expérimentation « Classification des masses mammaires basée sur les attributs statistiques du premier ordre »** : se focalise uniquement sur les mesures de texture du premier ordre, et pour tester leurs efficacités, trois classifieurs sont utilisés (RF, CVC et BN), la meilleure précision obtenue est de **76.8%** en utilisant les classifieurs « **Random Forest** » et « **Bayes network** ».
- **La quatrième expérimentation « Représentation profonde de la texture via l'opérateur LBP »** : consiste à utiliser deux types d'attributs de texture (les caractéristiques de Haralick et l'opérateur LBP) afin de bien représenter la masse mammaire ; et pour vérifier la validité de la méthode proposée, six classifieurs ont été utilisés (CVC, BN, DT, DS, J48 et RT), le meilleur taux de classification atteint égal à **64.3%** en utilisant le classifieur « **Random Tree** ».

- **La cinquième expérimentation** « *La discrimination des masses mammaires par la super-résolution et les descripteurs de texture* » : qui représente notre contribution majeur ; l'un de ces points clés consiste à introduire la technique de « la super-résolution » afin de discriminer les tumeurs malignes des tumeurs bénignes ; l'autre point important est de choisir un ensemble de descripteurs qui ont une interprétation ç'est à dire « les descripteurs les plus forts » parmi ceux qui ont été déjà testés et appliqués à partir de différentes zones d'intérêt dans les images mammographiques, et parmi la grande variété des descripteurs de texture existante, nous avons choisi de combiner dix-huit descripteurs de texture du premier et du troisième ordre. Afin de visualiser les performances de notre approche, neuf techniques de classification sont utilisées (SC, DT, TS, RF, MCC, CVR, LWL, SL et Logistic), le meilleur taux de précision obtenu est **96,7 %** en utilisant le classifieur « **Simple Logistic** ».

Au vue des résultats satisfaisants atteints, les conclusions suivantes peuvent être dégagées :

- La fusion des descripteurs du premier ordre et de Haralick renforce le taux de classification correcte des masses mammaires.
- Les caractéristiques du premier ordre aident à différencier les cas de tumeurs malignes des cas bénins.
- L'utilisation de l'opérateur LBP et les descripteurs statistiques du deuxième ordre assure une bonne représentation de texture des masses.
- L'intégration de la technique de la super-résolution dans un système d'aide au diagnostic est très avantageuse, car elle augmente la chance de discrimination des masses mammaires.
- La combinaison des caractéristiques du premier et du troisième ordre en utilisant une reconstruction à super-résolution améliore la capacité de connaître les masses malignes des masses bénignes.
- Nos travaux ont prouvé que « les attributs statistiques et les attributs basés sur la modélisation spatiale » ont un rôle important dans la caractérisation des masses mammaires dans l'image mammographique et malgré leurs simplicités de calcul, ils sont capables d'augmenter le taux de la classification correcte.
- Dans notre dernière contribution, nous confirmons que le « Simple Logistic » est le classifieur le plus adapté et le plus efficace dans notre système d'aide au diagnostic, car il a amélioré considérablement les taux de précision, de sensibilité et de spécificité.

- La comparaison de notre approche proposée qui a fait l'objet de la production scientifique de (Boudraa et al., 2020) avec d'autres approches de l'état de l'art a montré que nos résultats obtenus sont promoteurs et elle a confirmé l'efficacité de l'approche proposée ; c'est pourquoi, il serait intéressant d'appliquer notre approche sur d'autres modalités d'images médicales.
- L'exploitation de notre système d'aide au diagnostic par des experts sera favorable.

3. Perspectives

Mis à part nos bons résultats atteints, l'amélioration du taux de la classification correcte est nécessaire, c'est pourquoi les directions futures suivantes peuvent être évoquées :

- Il est intéressant d'améliorer la performance de notre système proposé avec d'autres travaux futurs, car la notion de texture qui est une donnée clé nécessite toujours le bon choix des descripteurs calculés afin d'avoir un taux de classification vraiment considérables et discriminer les zones cancéreuse des zones saine.
- L'exploitation d'autres bases de données sera avantageuse.
- Il est possible d'envisager d'autres voies de recherche, et pourquoi pas d'aller vers une hybridation de notre approche proposée avec une approche d'apprentissage profond.

Annexes

Annexe A « Attributs statistiques de texture »

➤ Attributs des histogrammes des sommes et des différences

L'ensemble des attributs statistiques d'ordre deux extraits des histogrammes des sommes et des différences sont (Porebski, 2009) :

$$\text{Centre de gravité somme} = \sum_{i=0}^{2(N-1)} i * P_S(i) = us \quad (43)$$

$$\text{Variance-somme} = \sum_{i=0}^{2(N-1)} (i - us)^2 P_S(i) \quad (44)$$

$$\text{Cluster shade-somme} = \sum_{i=0}^{2(N-1)} (i - us)^3 P_S(i) \quad (45)$$

$$\text{Cluster prominence-somme} = \sum_{i=0}^{2(N-1)} (i - us)^4 P_S(i) \quad (46)$$

$$\text{Entropie-somme} = - \sum_{i=0}^{2(N-1)} P_S(i) \log P_S(i) \quad (47)$$

$$\text{Centre de gravité-différence} = \sum_{j=-(N-1)}^{N-1} j * P_D(j) = ud \quad (48)$$

$$\text{Variance-différence} = \sum_{j=-(N-1)}^{N-1} (j - ud)^2 P_D(j) \quad (49)$$

$$\text{Cluster shade- différence} = \sum_{j=-(N-1)}^{N-1} (j - ud)^3 P_D(j) \quad (50)$$

$$\text{Cluster prominence- différence} = \sum_{j=-(N-1)}^{N-1} (j - ud)^4 P_D(j) \quad (51)$$

$$\text{Entropie- différence} = \sum_{j=-(N-1)}^{N-1} P_D(j) \log P_D(j) \quad (52)$$

$$\text{Energie} = \sum_{i=0}^{2(N-1)} (P_S(i))^2 * \sum_{j=-(N-1)}^{N-1} (P_D(j))^2 \quad (53)$$

$$\text{Contraste} = \sum_{i=0}^{N-1} \sum_{j=-(N-1)}^{N-1} j^2 P_D(j) \quad (54)$$

$$\text{Corrélation} = \sum_{i=0}^{2(N-1)} (i - us)^2 P_S(i) - j^2 P_D(j) \quad (55)$$

$$\text{Variance} = \text{Variance} - \text{somme} + \text{Contraste} \quad (56)$$

$$\text{Homogénéité} = \sum_{i=0}^{2(N-1)} \sum_{j=-(N-1)}^{N-1} \frac{1}{1+j^2} P_D(j) \quad (57)$$

$$\text{Entropie} = \text{Entropie-somme} + \text{Entropie- différence} \quad (58)$$

Où P_S : Histogrammes des sommes

P_D : Histogrammes des différences

➤ **Attributs de la matrice de Gray Levels Size Zone Matrix (GLSZM)**

L'ensemble des attributs statistiques d'ordre supérieur extraits de la matrice de Gray Levels Size Zone Matrix (GLSZM) sont (Mi et al., 2015):

$$\text{Small Zone Emphasis (SZE)} = \sum_i^M \sum_j^N \frac{P(i,j)}{j^2} \quad (59)$$

$$\text{Large Zone Emphasis (LZE)} = \sum_i^M \sum_j^N j^2 P(i, j) \quad (60)$$

$$\text{Low Gray Level Zone Emphasis (LGLZE)} = \sum_i^M \sum_j^N \frac{P(i,j)}{i^2} \quad (61)$$

$$\text{High Gray Level Zone Emphasis (HGLZE)} = \sum_i^M \sum_j^N i^2 P(i, j) \quad (62)$$

$$\text{Small Zone Low Gray Level Emphasis (SZLGLE)} = \sum_i^M \sum_j^N \frac{P(i,j)}{i^2 j^2} \quad (63)$$

$$\text{Small Zone High Gray Level Emphasis (SZHGLE)} = \sum_i^M \sum_j^N \frac{i^2 P(i,j)}{j^2} \quad (64)$$

$$\text{Large Zone High Gray Level Emphasis (LZLGLE)} = \sum_i^M \sum_j^N \frac{j^2 P(i,j)}{i^2} \quad (65)$$

$$\text{Large Zone Low Gray Level Emphasis (LZHGLE)} = \sum_i^M \sum_j^N i^2 j^2 P(i, j) \quad (66)$$

$$\text{Gray Level Non- Uniformity (GLNU)} = \sum_i^M \left(\sum_j^N P(i, j) \right)^2 \quad (67)$$

$$\text{Zone Size Non- Uniformity (ZSNU)} = \sum_j^N \left(\sum_i^M P(i, j) \right)^2 \quad (68)$$

$$\text{Zone Percentage (ZP)} = \frac{R}{jP(i,j)} \quad (69)$$

Où :

P (i; j) l'élément de coordonnées (i, j) de la matrice.

R : la somme des P (i,j)

Annexe B « Les algorithmes de classification avec Weka »

❖ Les classifieurs :

Comme nous avons déjà mentionné précédemment, WEKA est un puissant outil d'exploration de données, il inclut plus de 60 algorithmes d'apprentissage (Armutlu et al., 2008) ; Les classifieurs sont classés en sept classes (Bayes, Trees, Rules, Functions, Meta, Lazy, Misc) (Witten and Frank, 2002 ; Frank et al., 2016) (voir tableau B.1)

Classe	Classifieur
Bayes	AODE (<i>Averaged, one-dependence estimators</i>) Bayes Net Complement Naive Bayes Naive Bayes Naive Bayes Multinomial Naive Bayes Simple Naive Bayes Updateable
Trees	AD-Tree (<i>alternating decision trees</i>) Decision Stump Id3 J48 LMT M5P NB-Tree Random Tree RandomForest REP Tree Simple Cart User Classifier
Rules	ConjunctiveRule Decision Table JRip M5Rules Nnge (<i>Nearest-neighbor method of generating rules</i>) OneR Part Prism

	Ridor ZeroR
Functions	Least-MedSq Linear Regression Logistic Multilayer Perceptron Pace Regression RBF-Network Simple Linear Regression Simple Logistic SMO (<i>Sequential minimal optimization</i>) SMOreg (<i>Sequential minimal optimization algorithm for support vector regression</i>) Voted Perceptron Winnow
Lazy	IB1 (<i>Basic nearest-neighbor instance</i>) Ibk (<i>k-nearest-neighbor classifier</i>) Kstar (<i>Nearest neighbor with generalized distance function</i>) LBR (<i>Lazy Bayesian Rules</i>) LWL (<i>locally weighted learning</i>)
Meta	AdaBoostM1 Additive Regression Attribute selected classifier Bagging CVR (<i>Classification-via-regression</i>) Cost Sensitive Classifier CVParameter Selection CVC (<i>Classification-via-clustering</i>) Grading Logit Boost Meta cost Multi Boost AB Multi class classifier Rotation forest

	Threshold selector Vote
Misc	Hyperpipes VFI (<i>Voting feature intervals</i>)

Tableau B.1 : Les algorithmes de classification

Les classifieurs les plus utilisés sont :

➤ **Réseau bayésien (*Bayes network*)**

Le classifieur « réseau Bayésien » est parmi les techniques de classification les plus intéressantes de l'intelligence artificielle (Bouzaïeni, 2014).

Un réseau Bayésien est un modèle graphique probabiliste basé sur la théorie de Bayes (Vaithyanathan et al., 2013), où les nœuds représentent les variables aléatoires, liés par des arcs orientés et accompagnés de leurs indépendances conditionnelles (Bouzaïeni, 2014).

➤ **J48 ou Arbre de décision**

Les algorithmes d'arbres sont des méthodes simples aboutissant à des structures arborescentes (Marcellin, 2008) ; l'algorithme J48 est appelé implémentation optimisée du classifieur C4.5 ou version améliorée de la C4.5. La sortie donnée par le J48 est l'arbre de décision. Un arbre de décision est le même que celui de la structure arborescente ayant différents nœuds, tel que le nœud racine, les nœuds intermédiaires et le nœud feuille. Chaque nœud de l'arbre contient une décision et cette décision mène à un résultat qui s'appelle arbre de décision. L'arbre de décision divise l'espace d'entrée d'un ensemble de données en zones mutuellement exclusives, où chaque zone a une étiquette, une valeur ou une action pour décrire ou élaborer ses points de données (Nurul and Ahsan, 2015).

➤ **Souche de décision (*Decision stump*) :**

Une souche de décision est un arbre de décision, qui n'utilise qu'un seul attribut pour le fractionnement, pour les attributs discrets, cela signifie généralement que l'arbre n'est constitué que d'un seul nœud intérieur (c'est-à-dire que la racine n'a que des feuilles comme nœuds successeurs). Si l'attribut est numérique, l'arbre peut être plus complexe (Decision Stump, 2011).

➤ **Les arbres aléatoires (*Random Trees*) :**

L'algorithme des « arbres aléatoires » a été introduit par Leo Breiman et Adele Cutler. Il utilise une idée de mise en sac pour produire un ensemble de données aléatoires pour construire un arbre de décision. Dans un arbre standard, chaque nœud est divisé en utilisant la meilleure répartition parmi toutes les variables (Kalmegh, 2015).

Le concept du forêt aléatoire (*Random Forest*) est de construire des forêts aléatoires en mettant en sac des ensembles d'arbres aléatoires (Frank et al., 2016). Dans une forêt aléatoire (*random forest*), chaque nœud est divisé en utilisant le meilleur parmi le sous-ensemble de prédicteurs choisis au hasard à ce nœud. L'algorithme peut traiter à la fois les problèmes de classification et de régression.

Les arbres aléatoires sont une collection (ensemble) de prédicteurs d'arbres que l'on appelle forêt. La classification fonctionne comme suit : le classificateur d'arbres aléatoires prend le vecteur de caractéristique d'entrée, le classifie avec chaque arbre de la forêt, et produit le label de classe qui a reçu la majorité des "votes". En cas de régression, la réponse du classificateur est la moyenne des réponses sur tous les arbres de la forêt (Kalmegh, 2015).

➤ ***Les arbres de classification et de régression (Simple Cart) :***

Simple Cart (*Classification and Regression Tree*) est une méthode qui est utilisée pour l'exploration et la prédiction des données, elle utilise l'échantillon d'apprentissage, qui est un ensemble de données historiques avec des classes pour toutes les observations afin de construire l'arbre de décision.

Simple Cart génère l'arbre de décision binaire, et comme la sortie est un arbre binaire, elle ne génère que deux enfants. L'entropie est utilisée pour choisir le meilleur attribut de fractionnement (Kalmegh, 2015).

➤ ***Table de décision (Decision table)***

La table de décision est une méthode précise de prédiction numérique à partir d'arbres de décision et c'est un ensemble ordonné de règles "Si-Alors" qui ont le potentiel d'être plus compactes et donc plus compréhensibles que les arbres de décision (Ghogare, 2019). L'évaluation des sous-ensembles de caractéristiques est basée sur la recherche du meilleur premier et la validation croisée peut être utilisée (Frank et al., 2016).

C'est un algorithme plus simple et moins intensif en calcul que l'approche basée sur les arbres de décision (Ghogare, 2019).

➤ ***Simple logistic***

Le classificateur Simple logistic construit des modèles de régression logistique linéaire basés sur un seul attribut (Armutlu et al., 2008), en les adaptant à l'aide de l'algorithme de boosting « LogitBoost » et avec des fonctions de régression simple, en déterminant le nombre d'itérations à effectuer grâce à la validation croisée - qui prend en charge la sélection automatique des attributs (Frank et al., 2016).

➤ **Logistic**

Le classifieur **Logistic** construit un modèle de régression logistique à deux classes, il s'agit d'un modèle de régression statistique, où la régression logistique suppose que le rapport de vraisemblance logarithmique des distributions de classes est linéaire dans les observations (Armutlu et al., 2008).

➤ **LWL (locally weighted learning)**

LWL est un modèle d'apprentissage qui appartient à la catégorie des classifieurs basés sur la mémoire (Armutlu et al., 2008), il attribue des poids aux instances de formation en utilisant une fonction de pondération pour la classification ; Les poids sont calculés en deux étapes, la première est le calcul de la distance entre deux objets d'entrée et la deuxième est le calcul du poids pour chaque distance en utilisant une fonction noyau (*kernel function*). Tous les voisins sont utilisés pour déterminer la largeur de la fonction de pondération en utilisant le noyau linéaire, et une souche de décision est utilisée comme classificateur de base. Une méthode ANN linéaire est utilisée pour rechercher le plus proche voisin (Jha et al., 2019).

➤ **Classification par régression**

CVR « Classification-Via-Regression » effectue la classification en utilisant une méthode de régression (Frank et al., 2016). Lors de la classification, la classe dont le modèle donne la plus grande valeur de probabilité approximative est choisie comme classe prédite (Peckov, 2012).

➤ **La classification par regroupement**

La classification par regroupement CVC (Classification-Via-clustering) est un méta-classifieur simple qui utilise des algorithmes de regroupement, tels que ou les K-moyennes (*K-means*), les X-moyennes(*X-means*) et les méthodes de regroupement hiérarchique pour la classification (Jha et al., 2019).

➤ **Classifieur multi-classes (Multi-class classifier)**

Multi Class Classifier utilise quatre méthodes distinctes de classification à deux classes pour les problèmes multiclasse (Armutlu et al., 2008) en utilisant l'une de ces méthodes (Frank et al., 2016) :

- Un contre tous les autres.
- Classification par paires en utilisant le vote pour prédire.
- Codes exhaustifs de correction d'erreurs.
- Codes correcteurs d'erreurs sélectionnés au hasard.

➤ **Threshold selector**

Threshold Selector, optimise l'une des différentes mesures d'évaluation, notamment la mesure F, la précision, le rappel, l'exactitude et le taux positif réel, par une sélection d'un seuil de

probabilité sur la sortie du classificateur. Les performances peuvent être mesurées sur la base des données de formation par validation croisée (Frank et al., 2016).

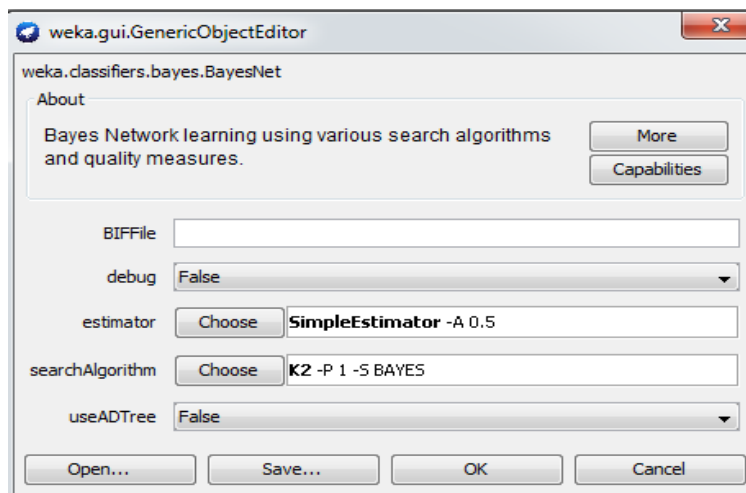
➤ **La forêt de rotation (Rotation forest)**

La forêt de rotation est un ensemble d'arbres qui effectue des transformations sur des sous-ensembles d'attributs avant de construire chaque arbre. Il y a deux principales différences entre la forêt aléatoire et la forêt de rotation (Bagnall et al., 2018):

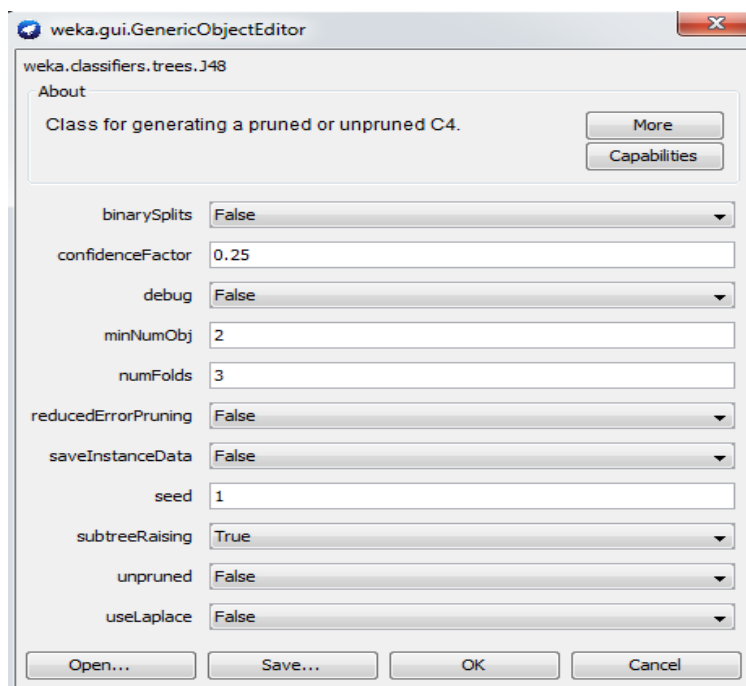
- La forêt de rotation transforme les attributs en ensembles
- La forêt de rotation utilise un arbre de décision C4.5

❖ **Les paramètres par défaut de quelques classifieurs dans Weka :**

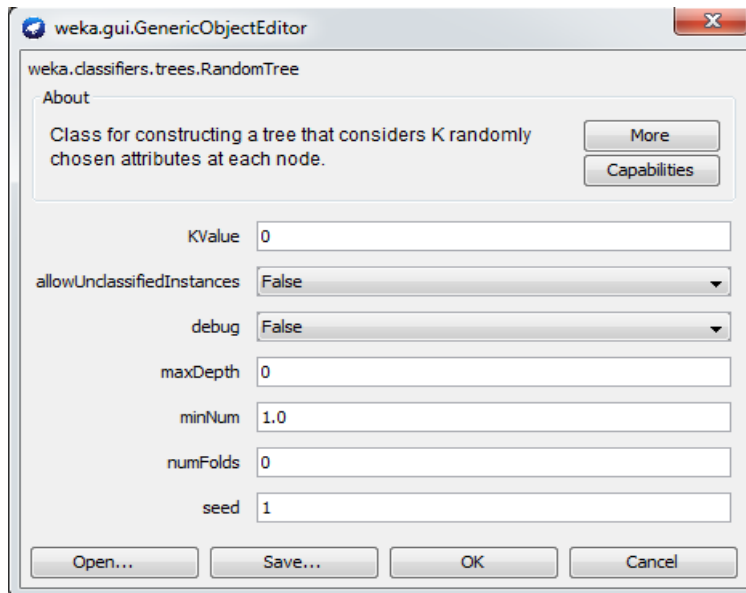
➤ **Bayes network**



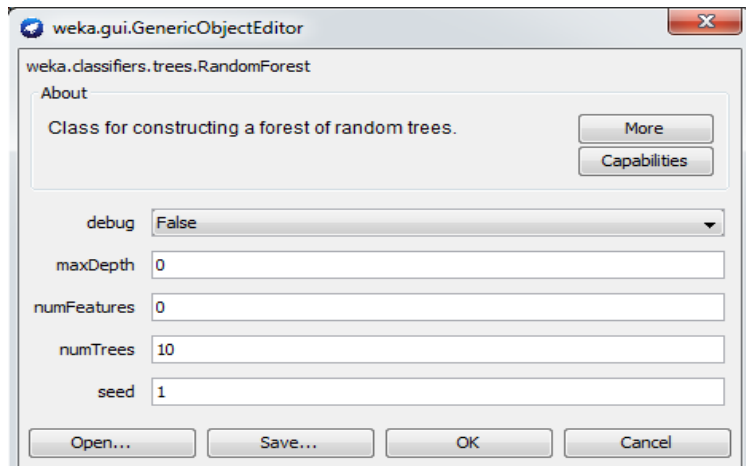
➤ **J48**



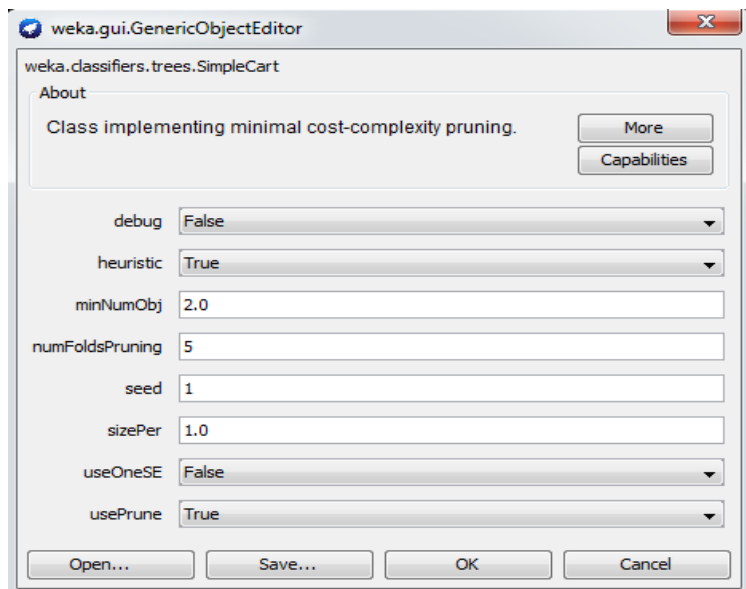
➤ *Random tree*



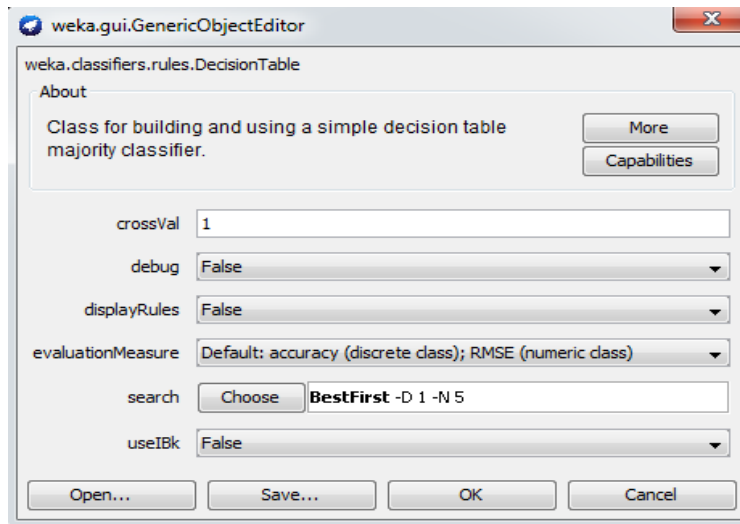
➤ *Random forest*



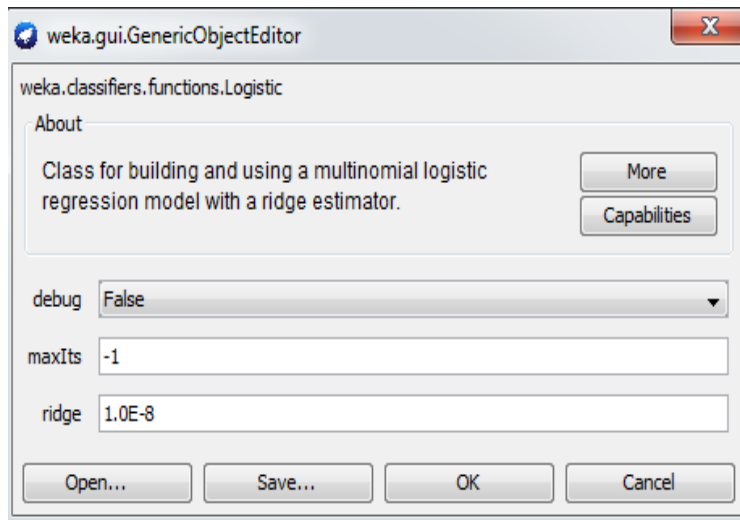
➤ *Simple cart*



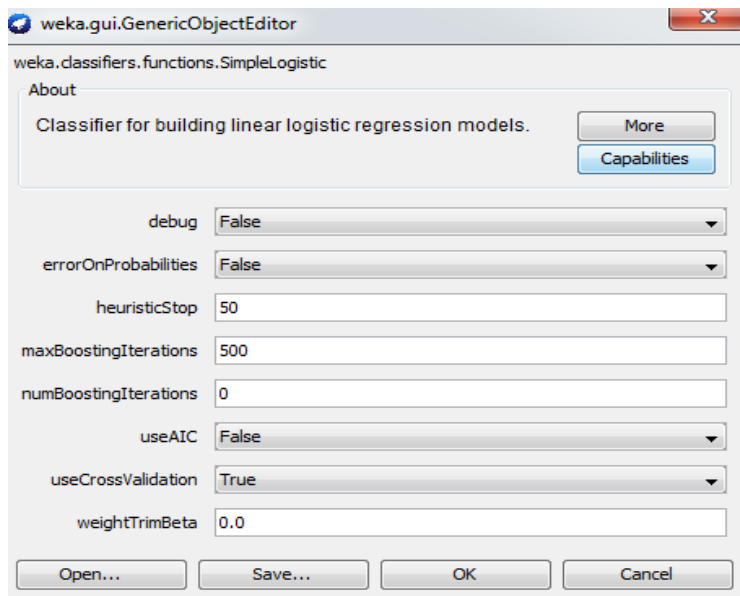
➤ *Decision table*



➤ *Logistic*



➤ *Simple logistic*



➤ **Multi Layer Perceptron**

The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.classifiers.functions.MultilayerPerceptron' classifier. The 'About' section describes it as 'A Classifier that uses backpropagation to classify instances.' Below this, various parameters are set:

- GUI: False
- autoBuild: True
- debug: False
- decay: False
- hiddenLayers: a
- learningRate: 0.3
- momentum: 0.2
- nominalToBinaryFilter: True
- normalizeAttributes: True
- normalizeNumericClass: True
- reset: True
- seed: 0
- trainingTime: 500
- validationSetSize: 0
- validationThreshold: 20

Buttons at the bottom include 'Open...', 'Save...', 'OK', and 'Cancel'.

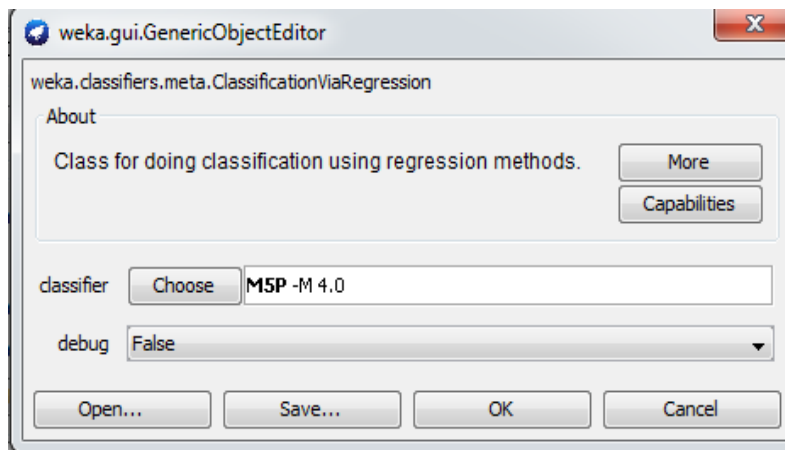
➤ **LWL**

The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.classifiers.lazy.LWL' classifier. The 'About' section describes it as 'Locally weighted learning.' The parameters are set as follows:

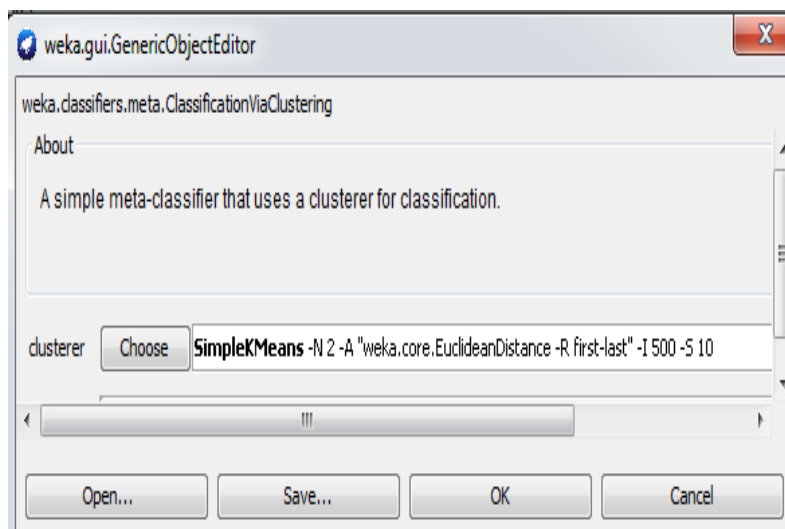
- KNN: -1
- classifier: Choose DecisionStump
- debug: False
- nearestNeighbourSearchAlgorithm: Choose LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"
- weightingKernel: 0

Buttons at the bottom include 'Open...', 'Save...', 'OK', and 'Cancel'.

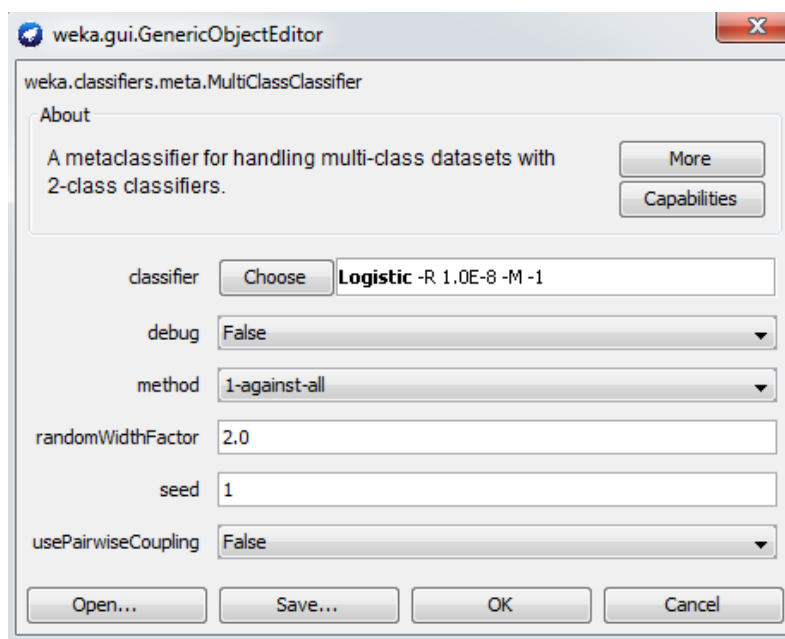
➤ *Classification-via-regression*



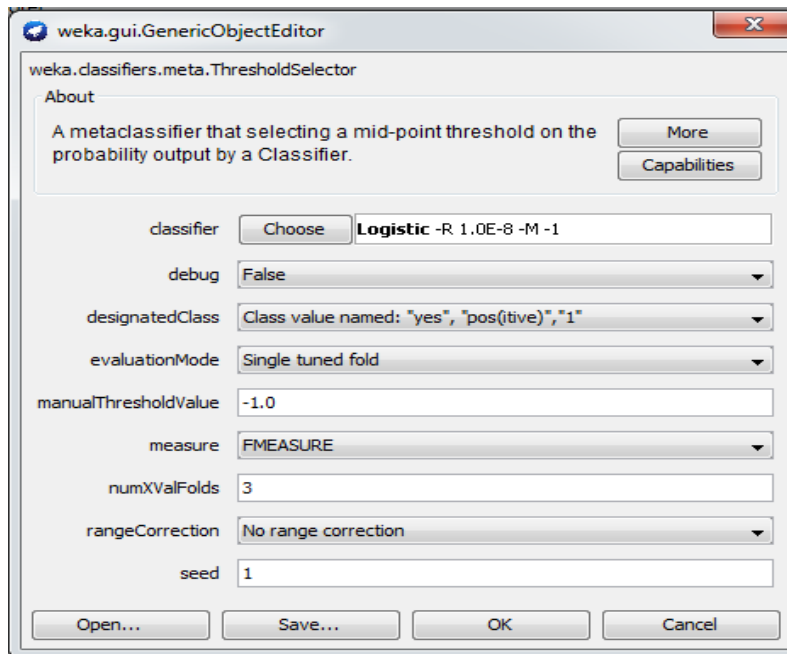
➤ *Classification-via-clustering*



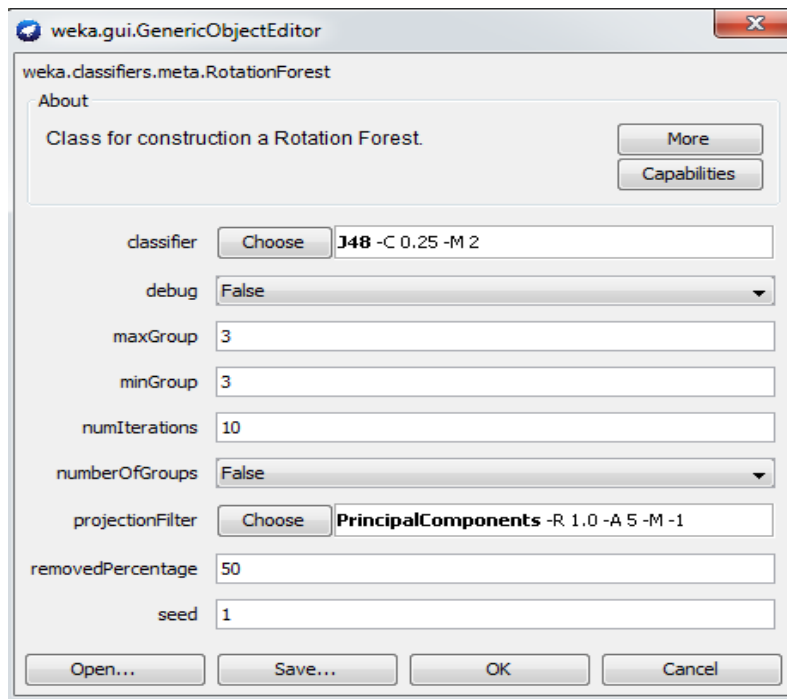
➤ *Multi-class classifier*



➤ *ThresholdSelector*



➤ *Rotation Forest*



Annexe C « Implémentation de l'approche proposée »

Dans cette partie, nous présentons des parties du code source de quelques classes de notre application, qui se base essentiellement sur l'extraction de la région d'intérêt et la classification, en utilisant la reconstruction avec la technique de la super-résolution.

➤ La classe « Calcul.java »

Cette classe englobe les fonctions qui calculent les caractéristiques de la l'image complète (la moyenne, la variance, l'aplatissement, l'uniformité, la variance, etc.)

```
*Haralick_propr.java  Calcul.java ⌵
35     }
36
37     private void cal_Moy_Var_ect()
38     {
39         for (int i = 0; i < b.length; i++) moy += b[i] ;
40         moy=moy/b.length;
41         for (int i = 0; i < b.length; i++) var += Math.pow((double)b[i]-moy, 2) ;
42         var= var/cpt;
43         ecartype=Math.sqrt(var);
44     }
45
46     /* la moyenne */
47     public double getmoy() { return moy; }
48
49     /*la variance*/
50     public double variance() { return var ;}
51
```

➤ La classe “Haralick_propr.java ”

Cette classe englobe les fonctions qui calculent les caractéristiques de la matrice “GLRLM” (Sum, SRE, LRE, GLNUu..., etc.)

```

542 /* ----- Calcul des caractéristiques GLRLM
543 public double Sum(double [][]matrix,int haut)
544 {double Sum = 0.0 ;
545 for (int y=0 ; y < 256 ; y++)
546     for (int x=0 ; x < haut ; x++)
547         Sum += matrix[y][x] ;
548 return Sum;}
549
550 //SRE
551 public double SRE(double [][]matrix ,int haut)
552 {
553     int n, l ;
554     double val = 0.0 ;
555     for (n=0 ; n < 256 ; n++)
556         for (l=0 ; l < haut ; l++)
557             val += matrix[n][l] / Math.pow(l, 2.0) ;
558     return val/ Sum(matrix, haut) ;
559 }
560 //LRE
561
562 public double LRE(double [][]matrix ,int haut)
563 {
564     int n, l ;
565     double val = 0.0 ;
566     for (n=0 ; n < 256 ; n++)
567         for (l=0 ; l < haut ; l++)
568             val += matrix[n][l] * Math.pow(l, 2.0) ;
569     return val / Sum(matrix, haut) ;
570 }
571
572 //GLN
573
574 public double GLNUu(double [][]matrix ,int haut)
575 {
576     int n, l ;
577     double v, val = 0.0 ;
578     for (n=0 ; n < 256 ; n++)
579         {

```

Dans cette partie, nous appelons toutes les fonctions qui calculent les caractéristiques du premier ordre (La moyenne, l'écart type, l'entropie, kurtosis, skewness, uniformity, regularity) et les caractéristique du troisième ordre (SRE, LRE, GLNun, RP, RLNu, LGLRE, HGLRE, SRHGLE et LRHGLE), ensuite, nous les rajoutons au vecteur de caractéristiques.

```
private void jButton5ActionPerformed(java.awt.event.ActionEvent evt) throws Exception {
    // Add to base TODO add your handling code here:
    Calcul cal=new Calcul(l, h, px);
    double mm=cal.getmoy(); double kk=cal.ecartype();double jj=cal.Kurtosis();
    double ll=cal.Skewness();
    double nn=cal.entropie();
    double qq=cal.Uniformity();
    double ww=cal.regularity();
    Haralick_propr hrlk=new Haralick_propr();
    m = new int[l * h];
    for (int k = 1; k < px.length; k++) {
        int r = px[k];
        int g = getBleu(r);
        m[k] = g;
    }
    Conversion conv = new Conversion(l, h, m);
    mat = conv.Vecteur_Matrice();
    Matrice_co_occurence mtrx=new Matrice_co_occurence(l,h,mat);
    ///AJOU CARAC LENGHT
    // LENGHT TODO add your handling code here:
    double [][] mcoo3 = new double [256][mat.length];
    mcoo3=mtrx.getMatrice1(mat);
    //mcoo=mtrx.mco();
    double secl=hrlk.SRE(mcoo3, h);
    double al=hrlk.LRE(mcoo3, h);
    double bl=hrlk.GLNUu(mcoo3, h);
    double cl=hrlk.RP(mcoo3, h);
    double dl=hrlk.RLNU(mcoo3, h);
    double el=hrlk.LGLRE(mcoo3, h);
    double fl=hrlk.HGLRE(mcoo3, h);
    double gl=hrlk.SRLGLE(mcoo3, h);
    double hl=hrlk.LRLGLE(mcoo3, h);
    double jl=hrlk.SRHGLE(mcoo3, h);
    double kl=hrlk.LRHGLE(mcoo3, h);
}
```

Ce qui concerne la reconstruction en utilisant la technique de « super-résolution », nous avons utilisé un plugin dans **ImageJ** qui s'appelle « **Parallel Super-Resolution** », il est disponible sur le site : <https://sites.google.com/site/piotrwendykier/software/parallelsuperresolution>.

Références bibliographiques

(Abdel-Nasser et al., 2016) Abdel-Nasser, M., Moreno, A. and Puig, D. (2016) 'Towards cost reduction of breast cancer diagnosis using mammography texture analysis', journal of experimental and theoretical artificial intelligence, vol. 28, pp. 385-402, doi : 10.1080/0952813X.2015.1024496.

(Abdel-Zaher and Eldeib, 2016) Abdel-Zaher, A. M. and Eldeib, A. M. (2016) 'Breast cancer classification using deep belief networks,' Expert Systems with Applications, vol. 46, pp. 139-144, doi : 10.1016/j.eswa.2015.10.015.

(Aboutalib et al., 2018) Aboutalib S. S., Abdelrahim, A., Berg, W. A., Zuley, M. L, Sumkin, J.H. and Wu, S. (2018) 'Deep Learning to Distinguish Recalled but Benign Mammography Images in Breast Cancer Screening', American Association for Cancer Research, vol. 24, doi : 10.1158/1078-0432.CCR-18-1115.

(Aggarwal and Aggarwal, 2012) Aggarwal, N. and Agrawal, R. K. (2012) 'First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images', Journal of Signal and Information Processing, vol. 3, pp. 146-153, doi : 10.4236/jsip.2012.32019.

(Alasadi and Baiee, 2013) Alasadi, T., Baiee, W. (2013) 'Classification of GIS Image using GLCM and Neural Network', Basrah Journal of Science, vol.31, pp. 110-119.

(Alharbi et al., 2012) Alharbi, H., Kwan, P., Jayawardena, A. and Sajeev, A., S., M. (2012) 'Fuzzy Image Segmentation for Mass Detection in Digital Mammography : Recent Advances and Techniques', Multidisciplinary Computational Intelligence Techniques : Applications in Business, Engineering, and Medicine, vol. 2, pp. 378-402, doi : 10.4018/978-1-4666-1830-5.ch021.

(Alsheh Ali, 2015) Alsheh Ali, M. (2015) 'Analyse statistique de populations pour l'interprétation d'images histologiques', thèse de doctorat, université Sorbonne Paris.

(Arevalo et al., 2016) Arevalo, J., Gonzalez, F. A., Ramos-Pollan, R., Oliveira, J. L., Angel, M. and Lopez, G. (2016) 'Representation learning for mammography mass lesion classification with convolutional neural networks', vol. 127, pp. 248-257, doi : 10.1016/j.cmpb.2015.12.014.

(Armutlu et al., 2008) Armutlu, P., Ozdemir ME, Uney-Yuksektepe F, Kavakli IH, Turkey M. (2008) 'Classification of drug molecules considering their IC50 values using mixed-integer

linear programming based hyper-boxes method', BMC Bioinformatics, vol. 9, doi : 10.1186/1471-2105-9-411.

(Attia, 2013) Attia, D. (2013) 'Segmentation d'images par combinaison adaptative couleur-texture et classification de pixels : Applications à la caractérisation de l'environnement de réception de signaux GNSS', thèse de doctorat, université de Technologie de Belfort-Montbeliard.

(Bagnall et al., 2018) Bagnall, A., Flynn, M., Large, J., Line, J., Bostrom, A. and Cawley, G. (2018) 'Is rotation forest the best classifier for problems with continuous features ?', arXiv : 1809.06705.

(Baratloo et al., 2015) Baratloo, A., Hosseini, M., Negida, A. and El Ashal, G. (2015) 'Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity', Emergency, vol. 3, pp. 48-49.

(Berbar, 2018) Berbar, M. A. (2018) 'Hybrid methods for feature extraction for breast masses classification', Egyptian Informatics Journal, vol. 19, pp. 63-73, doi : 10.1016/j.eij.2017.08.001.

(Bhanumathi and Suresh, 2015) Bhanumathi, R. and Suresh, G. R. (2015) 'Combining trace transform and SVN for classification of micro-calcifications in digital mammograms', 2nd international conference on electronics and communication system, India, pp. 1381-1386, doi : 10.1109/ECS.2015.7124811.

(Biermann, 2019) Biermann, J. (2019) 'Tumour evolution and novel biomarkers in breast cancer', thèse de doctorat, université de Göteborg.

(Boudraa et al., 2017a) Boudraa, S., Melouah, A. and Merouani, H. F. (2017) 'Une nouvelle combinaison de caractéristiques pour la classification de masses mammaires', Actes des 24èmes Rencontres de la Société Francophone de Classification, pp. 41-44, 28-30 juin, Lyon, France.

(Boudraa et al., 2017b) Boudraa, S., Melouah, A. and Merouani, H. F. (2017) 'Classification of mammography image using a new combination of statistics features', International Conference on Automatic control, Telecommunications and Signals, 11-12 Décembre, 2017, Annaba, Algérie.

(Boudraa et al., 2017c) Boudraa, S., Melouah, A. and Merouani, H. F. (2017) 'Statistics approach for breast mass classification', 3rd International Conference on Electrical

Engineering and Control Applications, 21-23 Novembre, Constantine, Algérie.

(Boudraa et al., 2018) Boudraa, S., Melouah, A. and Merouani, H. F. (2018) 'Deep texture representation for breast mass classification', 4th IEEE International Conference on Signal, Image, Vision and their Applications, 26-27 Novembre, Guelma, Algérie, pp. 1-4, doi : 10.1109/SIVA.2018.8661052.

(Boudraa et al., 2020) Boudraa, S., Melouah, A. and Merouani, H. F. (2020) 'Improving mass discrimination in mammogram-CAD system using texture information and super-resolution reconstruction', *Evolving Systems*, vol.11, pp. 697-706, doi : 10.1007/s12530-019-09322-4.

(Bouzaieni, 2014) Bouzaieni, A. (2014) 'Réseaux Bayésiens et quelques applications en traitement d'images', *RJC (17èmes Rencontres Jeunes Chercheurs)*, Nancy, 18-21 mars 2014, pp. 377-380, doi : 10.13140/2.1.2650.4000.

(Bovis and Singh, 2000) Bovis, K. and Singh, S. (2000) 'Detection of Masses in Mammograms Using Texture Features', *Proceedings 15th International Conference on Pattern Recognition*, Barcelona, vol. 2, pp. 267-270, doi : 10.1109/ICPR.2000.906064 .

(Bozek et al., 2009) Bozek J., Mustra M., Delac K. and Grgic M. (2009) 'A Survey of Image Processing Algorithms in Digital Mammography', In : Grgic M., Delac K. and Ghanbari M. (eds) *Recent Advances in Multimedia Signal Processing and Communications*, *Studies in Computational Intelligence*, vol. 231. Springer, Berlin, Heidelberg, doi : 10.1007/978-3-642-02900-4_24.

(Bray et al., 2018) Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018) 'Global cancer statistics 2018 : GLOBOCAN estimates of incidence and mortality world wide for 36 cancers in 185 countries', *CA Cancer J Clin*, vol. 68, pp. 394-424, doi : 10.3322/caac.21492.

(Braz et al., 2018) Braz, J.G., da Rocha, S.V., de Almeida, J.D.,S., de Paival, A. C., Silva, A. C. and Gattass M. (2018) 'Breast cancer detection in mammography using spatial diversity, geostatistics, and concave geometry', *Multimed Tools Appl*, vol. 78, pp. 13005-13031, doi : 10.1007/s11042-018-6259-z.

(Caroline and Vaijyanthi, 2016) Caroline, B. E. and Vaijyanthi, N. (2016) 'Feature extraction of digital mammogram based on multidimensional complete ensemble empirical mode decomposition with adaptive noise', *International Journal of Advanced Engineering Technology*, vol. VII, pp. 404-409.

- (Cheikhrouhou, 2012) Cheikhrouhou, I. (2012) 'Description and classification des masses mammaires pour le diagnostic du cancer du sein', thèse de doctorat, université d'évry val d'essonne.
- (Christensen-Jeffries et al., 2020) Christensen-Jeffries, K., Couture, O., Dayton, P. A., Eldar, Y. C., Hynynen, K., Kiessling, H., O'Reilly, M., Pinton, G. F., Schmitz, G., Tang, M., Tanter, M. and Sloun, R. J. G. (2020) 'Super-resolution Ultrasound Imaging', *Ultrasound in Medicine and Biology*, vol. 46, pp. 865-891, doi : 10.1016/j.ultrasmedbio.2019.11.013.
- (Chung and Nagy, 2008) Chung J. and Nagy, J. G. (2008) 'Nonlinear least squares and super resolution', *Journal of Physics : Conference Series*, vol. 124, doi : 10.1088/1742-6596/124/1/012019.
- (Cooper and Hausman, 2000) Cooper, G., and Hausman, R. E. (2000). 'The cell : a molecular approach', 2nd edition, Oxford : Sinauer Associates.
- (Daoudi and Khalifa, 2016) Daoudi, R. and Khalifa, D. (2016) 'Breast Cancer Classification by Artificial Immune Algorithm based Validity Interval Cells Selection', In *Proceedings of the 8th International Joint Conference on Computational Intelligence*, Porto, Portugal, pp. 209-216, doi : 10.5220/0006057202090216.
- (Datta et al., 2008) Datta, R., Joshi, D., Li, j. and Wang, J. Z. (2008) 'Image Retrieval : Ideas, Influences, and Trends of the New Age', *ACM Computing Surveys*, vol. 40, doi : 10.1145/1348246.1348248.
- (Decision Stump, 2011) Decision Stump (2011), In : Sammut C., Webb, G., I. (eds) 'Encyclopedia of Machine Learning', Springer, Boston, MA, doi : 10.1007/978-0-387-30164-8_202.
- (Delacour et al., 2005) Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J. F. and Ramirez, J. M., (2005) 'La courbe ROC (receiver operating characteristic) : principes et principales applications en biologie clinique', *Annales de Biologie Clinique*, vol. 63, pp. 145-154.
- (Dheeba et al., 2014) Dheeba, J., Singh, N. A. and Selvi, S. T. (2014) 'Computer-aided detection of breast cancer on mammograms : A swarm intelligence optimized wavelet neural network approach', *Journal of Biomedical Informatics*, vol. 49, pp. 45-52.
- (Dhungel et al., 2017) Dhungel, N., Carneiro, G. and Bradley, A. P. (2017) 'A Deep Learning Approach for the Analysis of Masses in Mammograms with Minimal User Intervention',

Medical Image Analysis, vol. 37, pp. 114-128.

(Doane et Lori 2011) Doane, D. P. and Lori, E. S. (2011) ‘Measuring Skewness: A Forgotten Statistic?’, Journal of Statistics Education, vol. 19, pp. 1-18.

(Dong et al., 2015) Dong, M., Lu, X., Ma, Y., Guo, Y., Ma, Y. and Wang, K. (2015) ‘An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms’, J Digit Imaging, vol. 28, pp. 613-625.

(Duraismy and Emperumal, 2017) Duraismy S. and Emperumal S. (2017) ‘Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier’, IAND Computer Vision, vol. 11, pp. 656-662.

(Eberl et al., 2006) Eberl, M. M, Fox, C. H., Edge, S. B., Carter, C. A. and Mahoney, M. C. (2006) ‘BI-RADS classification for management of abnormal mammograms’, J Am Board Fam Med, vol. 19, pp. 161-164, doi : 10.3122/jabfm.19.2.161.

(Eddaoudi, 2012) Eddaoudi, F., (2012) ‘aide à la détection des pathologies du sein par les techniques de segmentation et de codage de textures des images Mammographiques’, thèse de doctorat, université mohammed V– Agdal.

(El Atlas et al., 2014) El Atlas, N., El Aroussi, M. and Wahbi, M. (2014) ‘Computer-aided breast cancer detection using mammograms : A review’, Paper presented at the Second World Conference on Complex Systems.

(Farooqui and Ritkia, 2018) Farooqui, N. A. and Ritkia, M. (2018) ‘A study on early prevention and detection of breast cancer using three-machine learning techniques’, International Journal of Advanced Research in Computer Science U6 - Journal Article, vol. 9, pp. 37-42, doi : 10.26483/ijarcs.v9i0.6134.

(Fawcett, 2006) Fawcett, T. (2006) ‘An introduction to ROC analysis’, Pattern Recognition Letter, vol. 27, pp. 861-874.

(Ferlay et al., 2015) Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., M., Forman, D. and Bray, F. (2015) ‘Cancer incidence and mortality worldwide : sources, methods and major patterns in GLOBOCAN 2012’, Int J Cancer, vol 136, pp E359-86, doi : 10.1002/ijc.29210.

(Ferlay et al., 2019) Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D., Piñeros, M., Znaor, A. and Bray, F. (2019) ‘Estimating the global cancer incidence and

mortality in 2018 : GLOBOCAN sources and methods', *International Journal of Cancer*, vol. 144, pp. 1941-1953, doi : 10.1002/ijc.31937.

(Filipczuk et al., 2012) Filipczuk, P., Fevens, T., Krzyżak, A. and Obuchowicz, A. (2012). GLCM and GLRLM based texture features for computer-aided breast cancer diagnosis. *Journal of Medical Informatics and Technology*, vol. 19, pp. 109-115.

(Frank et al., 2016) Frank, E., Hall, M. A. and Witten, H. (2016), *The WEKA Workbench. Online Appendix for "Data Mining : Practical Machine Learning Tools and Techniques"*, (Fourth Edition), Morgan Kaufmann.

(Galloway, 1975) Galloway, M. M. (1975) 'Texture analysis using gray level run lengths', *Computer Graphics And Image Processing*, vol. 4, pp. 172-179.

(Garnier, 2014) Garnier, M. (2014) 'Modèles descriptifs de relations spatiales pour l'aide au diagnostic d'images biomédicales', Thèse de doctorat, université René Descartes - Paris V.

(Ghogare, 2019) Ghogare, S. A. (2019) 'Performance Comparison of Tree based classifier using WEKA', *journal of gujarat research society*, vol. 21 pp. 70-75.

(Görgel et al., 2015) Görgel, P., Sertbas, A., Uçan, O.N. (2015) 'Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines', *Expert Systems*, vol. 32, pp. 155-164.

(Greenspan, 2009) Greenspan, H. (2009) 'Super-resolution in medical imaging', *the computer journal*, vol. 52, pp. 43-63, doi : 10.1093/comjnl/bxm075.

(Grønvold, 2010) Grønvold, M. (2010) 'Health-related quality of life in early breast cancer', thèse de doctorat, *Danish Medical Bulletin*, vol.57.

(Guerra et al., 2011) Guerra, L., McGarry, L. M., Robles, V., Bielza, C., Larrañaga, P. and Yuste, R. (2011) 'Comparison between supervised and unsupervised classifications of neuronal cell types : a case study', *Dev Neurobiol.* vol. 71, pp. 71-82, doi : 10.1002/dneu.20809.

(Guliato et al., 2009) Guliato, D., Bôaventura, R. S., Maia, M. A., Rangayyan, R. M., Smedo, M. S. and Macedo, T. A. (2009) 'INDIAM--an e-learning system for the interpretation of mammograms', *Journal of digital imaging*, vol. 22, pp. 405-420, doi : 10.1007/s10278-008-9111-6.

(Hamoud et al., 2015) Hamoud M., Merouani H. F., Laimeche L. (2015) 'The power laws :

Zipf and inverse Zipf for automated segmentation and classification of masses within mammograms' .Evolving Systems, vol. 6, pp. 209-227.

(Haralick et al., 1973) Haralick R. M., Shanmugam K., Dinstein I. (1973) 'Textural features for image classification', IEEE transactions on systems, man, and cybernetics, vol. 3, pp. 610-21, doi : 10.1109/TSMC.1973.4309314.

(Hastie et al., 2008) Hastie, T., Tibshirani, R. and Friedman, J. (2008) 'The elements of statistical learning-data mining', inference and prediction, 2nd Edition ed., vol. II, Stanford : Springer, pp. 465-576.

(Hemmatian and Sohrabi, 2019) Hemmatian, F., Sohrabi, M. K. (2019) 'A survey on classification techniques for opinion mining and sentiment analysis', Artificial Intelligence Review, vol. 52, pp. 1495-1545, doi : 10.1007/s10462-017-9599-6.

(Herwanto and Arymurthy, 2013) Herwanto, Arymurthy, A. M. (2013) 'Association technique based on classification for classifying microcalcification and mass in mammogram' 'International Journal of Computer Science Issues, vol. 10, pp. 252-259.

(Hou et al., 2020) Hou, J., Si, Y. and Yu, X. (2020) 'A Novel and Effective Image Super-Resolution Reconstruction Technique via Fast Global and Local Residual Learning Model', Applied Sciences, vol. 10, pp. 1856-1867, doi : 10.3390/app10051856.

(Houam, 2013) Houam, L. (2013) 'Contribution à l'analyse de textures de radiographies osseuses pour le diagnostic précoce de l'ostéoporose', Thèse de doctorat, université d'Orléans.

(Jadoon et al., 2017) Jadoon, M. M., Zhang, Q., Haq, I. U., Butt, S. and Jadoon, A. (2017) 'Three-Class mammogram classification based on descriptive CNN features', Hindawi Biomed Research International, doi : 10.1155/2017/3640901.

(Jalalian et al., 2013) Jalalian, A., Mashohor, S. B., Mahmud, H. R., Saripan, M. I., Ramli, A. R. and Karasfi, B. (2013) 'Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound : a review', Elsevier Clinical Imaging, vol. 37, pp. 420-426, doi : 10.1016/j.clinimag.2012.09.024.

(Jalalian et al., 2017) Jalalian, A., Mashohor, S., Mahmud, R., Karasfi, B., Iqbal, M., Saripan, B. and Ramli, A.B. (2017) 'Foundation And Methodologies In Computer-Aided Diagnosis Systems For Breast Cancer Detection', EXCLI Journal, vol. 16, pp. 113-137, doi : 10.17179/excli2016-701.

- (Jasmine et al., 2013) Leena Jasmine, J.S., Baskaran, S. and Govardhan, A. (2013) 'A robust approach to classify microcalcification in digital mammograms using contourland transform and support vector machine', *American Journal of Engineering and Applied Sciences*, vol. 6, pp. 57-68.
- (Jehloul et al., 2015) Jehloul, H. B. khyioon A. Z., Subhi, A, and Oleiwi, A. (2015) 'classification of mammography image using machine learning classifiers and texture features', *International Journal of Innovative Research in Advanced Engineering*, vol. 2, pp. 56-63.
- (Jha et al., 2019) Jha, S. K., Pan, Z., Elahi, E, Patel, N. A. (2019) 'comprehensive search for expert classification methods in disease diagnosis and prediction', *Expert Systems*, vol. 36, doi : 10.1111/exsy.12343.
- (Jiao et al., 2016) Jiao, Z. Gao, X. Wang, Y. and Li, J. (2016) 'A deep feature based framework for breast masses classification', *Neurocomputing*, vol. 197, pp. 221-231.
- (Jinan, 2015) Jinan, F.M. (2015) 'Application of GLCM technique on Mammograms for Early Detection of Breast Cancer', *Journal of Babylon University/Pure and Applied Sciences*, vol. 23, pp. 885-890, 2015.
- (Kalmegh, 2015) Kalmegh, S. (2015) 'Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News', *International Journal of Innovative Science, Engineering & Technology*, vol. 2, pp. 438-446.
- (Kennedy et al., 2006) Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R. and Azhari, H. (2006) 'Super-resolution in PET imaging', *IEEE Transactions on Medical Imaging*, vol. 25, pp.137-147.
- (Kennedy et al., 2007) Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R. and Azhari, H. (2007) 'Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution', *International Journal of Biomedical Imaging*, doi : 10.1155/2007/46846.
- (Khan et al., 2017) Khan, S., Hussain, M., Aboalsamh, H. and Bebis, G. (2017) 'A comparison of different Gabor feature extraction approaches for mass classification in mammography', *Multimedia Tools and Application*, vol. 76, pp. 33-57.
- (Khan et al.,2016) Khan, A., Hussain, M., Aboalsamh, H., Mathkour, H., Bebis, G. and Zakariah, M. (2016) 'Optimized Gabor features for mass classification in mammography',

Applied Soft Computing, vol. 44, pp. 267-280.

(Kharman-Biz, 2016) Kharman-Biz, A. (2016) 'Potential novel molecular targets for breast cancer diagnosis and treatment', thèse de doctorat, Institut de Karolinska, Stockholm, Suède.

(Kraus et al., 2016) Kraus, O.Z, Ba, J.L, Frey, B.J. (2016) 'Classifying and segmenting microscopy images with deep multiple instance learning', Bioinformatics, vol. 32, pp. 52-i59.

(Krig, 2014) Krig, S. (2014) 'Computer vision metrics: survey, taxonomy, and analysis', Apress, USA, doi : 10.1007/978-1-4302-5930-5

(Lee et al., 2017) Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M. and Rubin, D. L. (2017) 'A curated mammography data set for use in computer-aided detection and diagnosis research', Scientific data, vol. 4, doi : 10.1038/sdata.2017.177.

(LeGal et al., 1984) LeGal, M., Chavanne, G., and Pellier, D. (1984) 'Valeur diagnostique des microcalcifications groupées d'écouvertes par mammographie. à propos de 227 cas', Bull Cancer, vol. 71, pp. 57-64.

(Lévy and Jain, 2016) Lévy, D. and Jain, A. (2016) 'Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks', arXiv Preprint. arXiv : 1612.00542.

(Li et al., 2001) Li, H., Wang, Y., Ray Liu, K.J., Lo, S.C. and Freedman, M.T. (2001) 'Computerized radiographic mass detection-part II : decision support by featured database visualization and modular neural networks', IEEE Transactions on Medical Imaging, vol. 20, pp. 302-313, doi : 10.1109/42.921479.

(Li et al., 2009) Li, Z., Imai, J. and Kaneko, M. (2009) 'Facial-component-based bag of words and PHOG descriptor for facial expression recognition', IEEE International Conference on Systems, Man and Cybernetics, San Antonio, pp. 1353-1358, doi : 10.1109/ICSMC.2009.5346254.

(Li et al., 2016) Li, Y., Chen, H., Cao, L. and Ma., J. (2016) 'A Survey of Computer-aided Detection of Breast Cancer with Mammography', Journal of Health and Medical Informatics, vol. 7, doi : 10.4172/2157-7420.1000238.

(Li et al., 2019) Li, H., Zhuang, S., Li, D., Zhao,J. and Ma,Y. (2019) 'Benign and malignant classification of mammogram images based on deep learning', Biomedical Signal Processing and Control, vol. 51, pp. 347-354, doi : 10.1016/j.bspc.2019.02.017.

- (Liantoni et al., 2019) Liantoni, F., Perwira, R., Dwi Putri, L., Manurung, R., Kahar, M., Safitri, J., Muharlisiani, L., Chamidah, D., Ghofur, A., Kurniawan, S., Wityasminingsih, E., Yuniningsih., Susilo., Yusuf, M., Kurniawan, A., Sumitro., Sutjiatmo, B., Muna, N., Fatmasari, F., Nawas and Tari, N. (2019) ‘Watermelon classification using k-nearest neighbours based on first order statistics extraction’, *Journal of Physics: Conference Series*, 1st International Conference on Advance and Scientific Innovation 23–24 April 2018, Medan, Indonesia vol. 1175, doi :10.1088/1742-6596/1175/1/012114
- (Liu and Tang, 2014) Liu, X. and Tang, J. (2014) ‘Mass Classification in Mammograms Using Selected Geomandry and Texture Features, and a New SVM-Based Feature Selection Mandhod’, *IEEE Systems Journal*, vol. 8, pp. 910-920.
- (Mabrouk et al., 2019) Mabrouk, M., Afify, H. and Marzouk, S. (2019) ‘Fully automated computer-aided diagnosis system for micro calcifications cancer based on improved mammographic image techniques’, *Ain Shams Engineering Journal*, vol. 10, pp. 517-527, doi : 10.1016/j.asej.2019.01.009.
- (Mäenpää and Pietikäinen, 2004) Mäenpää, T. and Pietikäinen, M. (2004) ‘texture analysis with local binary patterns’, *Handbook of Pattern Recognition and Computer Vision*, pp 197-216, doi : 10.1142/9789812775320_0011
- (Makwana and Mehta, 2013) Rujul R. Makwana Nita D. Mehta (2013) ‘Survey on Single image Super Resolution Techniques’, *IOSR Journal of Electronics and Communication Engineering*, Vol. 5, pp. 23-33, doi : 10.9790/2834-0552333.
- (Malhotra et al., 2010) Malhotra, G.K., Zhao, X., Band, H. and Band, V. (2010) ‘Histological, molecular and functional subtypes of breast cancers’, *Cancer Biol Ther*, vol. 10, pp. 955-960, doi : 10.4161/cbt.10.10.13879.
- (Mallapragada et al., 2010) Mallapragada, P. K., Jin, R. and Jain, A. (2010) ‘Non-parametric Mixture Models for Clustering’, In : Hancock E.R., Wilson R.C., Windeatt T., Ulusoy I., Escolano F. (eds) *Structural, Syntactic, and Statistical Pattern Recognition, SSPR /SPR 2010*, *Lecture Notes in Computer Science*, vol. 6218. Springer, Berlin, Heidelberg, doi : 10.1007/978-3-642-14980-1_32.
- (Marcellin, 2008) Marcellin, S. (2008) ‘Arbres de décision en situation d’asymétrie’, thèse de doctorat, université Lumière Lyon II.

(Martinez, 2004) Martinez, J.M. (2004) 'MPEG-7 overview (version 10)', ISO/IEC JTC1/SC29/WG11 N6828, Palma de Mallorca, Espagne.

(Meunier, 2010) Meunier, E. (2010) 'Interrelations entre les protéines Rho and le récepteur des oestrogènes alpha dans des modèles de cancers mammaires', thèse de doctorat, université de Toulouse.

(Mi et al., 2015) Mi, H., Petitjean, C., Dubray, B., Vera, P. and Ruan, S. (2015) 'Robust feature selection to predict tumor treatment outcome', *Artificial Intelligence In Medicine*, vol. 64, pp. 195-204, doi : 10.1016/j.artmed.2015.07.002.

(Mohanty et al., 2011) Mohanty, A.K., Beberta, S. and Lenka, S.K. (2011) 'Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram', *International Journal of Engineering Research and Applications*, Vol 1, pp 687-693

(Momenimovahed and Salehiniya, 2019) Momenimovahed Z. and Salehiniya H. (2019) 'Epidemiological characteristics of and risk factors for breast cancer in the world', *Breast Cancer (Dove Med Press)*, vol. 11, pp. 151-164, doi : 10.2147/BCTT.S176070.

(Moreira et al., 2012) Moreira, I., C., Amaral, I. Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J., S. (2012) 'INbreast : Toward a Full-field Digital Mammographic Database', *Academic Radiology*, vol. 19, pp. 236-48, doi : 10.1016/j.acra.2011.09.014.

(Mussarat et al., 2013) Mussarat, Y., Muhammad, S. and Sajjad, M. (2013) 'Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques', *Research Journal of Recent Sciences*, vol. 2, pp. 88-98.

(Nasrollahi and Moeslund, 2014) Nasrollahi, K. and Moeslund, T. B. (2014) 'Super-resolution: A comprehensive survey', *Machine Vision, Applications*, vol. 25, pp. 1423-1468, doi : 10.1007/s00138-014-0623-4.

(Nelson et al., 2016) Nelson, H. D., Cantor, A., Humphrey, L., Fu, R., Pappas, M., Daeges, M. and Griffin J. (2016) 'Screening for Breast Cancer : A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation [Internet]', Rockville (MD) : Agency for Healthcare Research and Quality (US), Rapport N°. 14-05201-EF-1.

(Nithya and Santhi, 2011) Nithya, R. and Santhi, B. (2011) 'Comparative study on feature extraction method for breast cancer classification', *Journal of Theoretical and Applied*

Information Technology', vol. 33, pp. 220-226.

(Nurhayati et al., 2011) Nurhayati, O. D, Susanto, A., Widodo, T. S. and Tjokronagoro, M. (2011) 'Principal Component Analysis combined with First Order Statistical Method for Breast Thermal Images Classification', International Journal of Computer Science and Telecommunications, vol. 2, pp. 12–18.

(Nurul and Ahsan, 2015) Nurul, A. and Habib, A. (2015) 'Comparison of Different Classification Techniques Using WEKA for Hematological Data', American Journal of Engineering Research, vol. 4, pp. 55-61.

(Odille et al., 2015) Odille, F., Bustin, A., Chen, B., Vuissoz, P. A. and Felblinger, J. (2015) 'Motion-corrected, super-resolution reconstruction for high-resolution 3D cardiac cine MRI', International Conference on Medical Image Computing, in Lecture Notes in Computer Science, pp. 435-442, doi : 10.1007/978-3-319-24574-4_52.

(Oeffinger et al., 2015) Oeffinger, K. C., Fontham, Andzioni, R., Herzig A., Michaelson, J. S., Shih, Y. C., Walter, L. C., Church, T. R., Flowers, C. R., LaMonte, S. J., Wolf, A. M., DeSantis, C., Lortand-Tieulent, J., Andrews K., Manassaram-Baptiste, D., Saslow, D., Smith, R. A, Brawley, O. W. and Wender, R. (2015) 'Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society', JAMA vol. 314: pp. 1599-1614, doi : 10.1001/jama.2015.12783.

(Ojala and Pietikäinen, 2004) Ojala, T. and Pietikäinen, M. (2004) 'Texture Classification', Machine Vision and Media Processing Unit, University of Oulu, Finland, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OJALA1/texclas.htm.

(Ojala et al., 1996) Ojala, T., Piandikainen, M. and Harwood, D. (1996). 'A comparative study of texture measures with classification based on feature distributions', Pattern Recognition, vol. 29, pp. 51-59.

(Olaode et al., 2014) Olaode, A., Naghdy, G., and Todd, C. (2014). 'Unsupervised Classification of Images : A Review', International Journal of Image Processing, vol. 8, pp. 325-342.

(Oliveira et al., 2015) Oliveira, F.S.S., Filho, A.O.C., Silva, A.C., Paiva, A.C., Gattass, M. (2014) 'Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM', Computers in Biology and Medicine, vol. 57, pp. 42-53.

(Padma and Vijayalakshmi, 2016) Padma, P. Vijayalakshmi, D. (2016) 'Image Retrieval

Using WaveletBased Interactive Genetic Algorithm’, International Journal of Emerging Technology and Advanced Engineering, vol. 6, pp. 178–183.

(Pak et al., 2015) Pak, F., Kanan, H. and Alikhassi, A. (2015) ‘Breast cancer detection and classification in mammography based on Non-Subsampled Contourlet Transform (NSCT) and super resolution’, computer methods and programs in biomedicine, vol. 22, pp. 89-107, doi : 10.1016/j.cmpb.2015.06.009.

(Papakonstantinou, 2020) Papakonstantinou, A. (2020) ‘Studies Of Side Effects Related To Adjuvant Breast Cancer Regimens With Focus On Chemotherapy’, thèse de doctorat, Institut Karolinska, Stockholm, Suède.

(Park et al., 2003) Park, S. C., Park, M. K. and Kang, M. G. (2003) ‘Super-resolution image reconstruction : a technical overview’, IEEE Signal Processing Magazine, vol. 20, pp. 21-36, doi : 10.1109/MSP.2003.1203207

(Paul et al., 2015) Paul, S., Solanki, P. P., Shahi, U. P., & Srikrishna, S. (2015) ‘Epidemiological Study on Breast Cancer Associated Risk Factors and Screening Practices among Women in the Holy City of Varanasi, Uttar Pradesh, India’. Asian Pacific journal of cancer prevention : APJCP, 16(18), 8163–8171. <https://doi.org/10.7314/apjcp.2015.16.18.8163>

(Paulhac, 2009) Paulhac, L. (2009) “Outils and méthodes d’analyse d’images 3D texturées : application à la segmentation des images échographiques”, thèse de doctorat, université François Rabelais – Tours, France.

(Peckov, 2012) Peckov, A. (2012) ‘A machine learning approach to polynomial regression’, thèse de doctorat, Jozef stefan international postgraduate school.

(Pezeshki et al.,2019) Pezeshki, H., Rastgarpour, M., Sharifi, A. and Yazdani, S. (2019) ‘Extraction of spiculated parts of mammogram tumors to improve accuracy of classification’, Multimedia Tools and Applications, vol. 78, pp. 19979-20003.

(Pham, 2018) Pham, C., H. (2018) ‘Apprentissage profond pour la super-résolution et la segmentation d’images médicales’, thèse de doctorat, école nationale supérieure Mines-Télécom Atlantique Bretagne.

(Porebski, 2009) Porebski, A. (2009) ‘Sélection d’attributs de texture couleur pour la classification d’images. Application à l’identification de défauts sur les décors verriers imprimés par sérigraphie’, thèse de doctorat, université de Lille.

- (Prasad and Basha, 2018) Prasad, K. R. and Basha, M. (2018) 'Effective texture feature model for classification of mammogram images', *ARPN Journal of Engineering and Applied Sciences*, vol. 13, pp. 961-967.
- (Pratiwia et al., 2015) Pratiwia, M., Alexander, Harefaa, J. and Nandaa, S. (2015) 'Mammograms Classification using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network', *International Conference on Computer Science and Computational Intelligence*, IN *Procedia Computer Science*, vol. 59, pp. 83-91.
- (Puhalla et al., 2012) Puhalla, S., Bhattacharya, S., Davidson, N. E. (2012) 'Hormonal therapy in breast cancer : a model disease for the personalization of cancer care', *Molecular oncology*, vol. 6, pp. 222-236, doi : 10.1016/j.molonc.2012.02.003.
- (Rabidas et al., 2016) Rabidas, R., Midya, A., Chakraborty, J. and Arif, W. (2016) 'A study of different texture features based on local operator for benign-malignant mass classification', *International Conference On Advances In Computing and Communications*, In *Procedia Computer Science*, Cochin, India, vol. 93, pp. 389-395.
- (Raghavendraa et al., 2016) Raghavendraa, U., Acharyab, U. R., Fujitae, H. Gudigara, A., Tanb, J. H. and Chokkadia, S. (2016) 'Application of Gabor waveland and Locality Sensitive Discriminant Analysis for Automated identification of breast cancer using digitized mammogram images', *Applied Soft Computing*, vol. 46, pp. 151-161.
- (Ramadan, 2020) Ramadan, S. Z. (2020) 'Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms : A Review', *Journal of Healthcare Engineering*, 2020, pp. 1-21. doi : 10.1155/2020/9162464
- (Ramsay et al., 2005) Ramsay, D.T., Kent, J. C., Hartmann, R., A. and Hartman, P., E. (2005) 'Anatomy of the lactating human breast redefined with ultrasound imaging', *Journal of anatomy*, vol. 206, pp. 525-534.
- (Rangra and Bansal, 2014) Rangra, K. and Bansal, K., L. (2014) 'Comparative Study of Data Mining Tools', *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, pp. 216-223.
- (Ribli et al., 2018) Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I. (2018) 'Detecting and classifying lesions in mammograms with Deep Learning', *Sciece Report*, vol. 8, pp. 1-7.
- (Rosenfeld, 1976) Rosenfeld, A. (1976) 'Digital Picture Analysis', Springer-Verlag Berlin

Heidelberg, doi : 10.1007/3-540-07579-8.

(Rouhi et al., 2015) Rouhi, R., Jafari, M. and Kasaei, S. (2015) 'Benign and malignant breast tumors classification based on region growing and CNN segmentation', *Expert Systems with Applications*, vol. 42, pp. 990-1002, doi : 10.1016/j.eswa.2014.09.020.

(Sabu et al., 2012) Sabu, A. M., Ponraj, D. N. and Poongodi, (2012) 'Textural Features Based Breast Cancer Detection: A Survey', *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, pp. 1329-1334.

(Sage, 2012) Sage, D. (2012) 'Analyse d'image scientifique, le monde ImageJ', *Flash informatique*, pp. 3-7.

(Sakikala and Ezhilarassi, 2016) Sakikala, S. and Ezhilarassi, M. (2016) 'Combination of Mammographi Texture Feature Descriptors for improved breast cancer diagnosis', *Asian J.Inform.Technol*, vol. 15, pp. 4054-4062.

(Sampat et al., 2005) Sampat, M.P., Markey, M.K., and Bovik, A.C. (2005) 'Computer-aided detection and diagnosis in mammography', In the handbook of image and video processing, pp. 1195-1217, doi : 10.1016/b978-012119792-6/50130-3.

(Saraswathi and Srinivasan, 2016) Saraswathi, D. and Srinivasan,E. (2016) 'A high-sensitivity computer-aided system for detecting microcalcifications in digital mammograms using curvelet fractal texture features', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 5, pp. 263-273, doi : 10.1080/21681163.2015.1089793.

(Saraswathi et al., 2015) Saraswathi, D., Srinivasan, E. and Dharani, D. (2015) 'An efficient feature extraction technique for breast cancer diagnosis using curveland transform and swarm intelligence', *International Conference on Wireless Communications, Signal Processing and Networking, India*, pp. 441-445, doi : 10.1109/WiSPNAND.2016.7566172.

(Sathya and Annamma, 2013) Sathya, R. and Annamma, A. (2013) 'Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification', *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, doi : 10.14569/IJARAI.2013.020206.

(Saygılı, 2018) Saygılı, A. (2018) 'Classification and Diagnostic Prediction of Breast Cancers via Different Classifiers', *International Scientific and Vocational Studies Journal*, vol. 2, pp.

48-56.

(Seryasat and Haddadnia, 2017) Seryasat, O., R. and Haddadnia J. (2017) 'Assessment of a novel computer aided mass diagnosis system in mammograms', *Biomedical Research*, vol. 28, pp. 3129-3135.

(Shanthi and Murali, 2014) Shanthi., S. and Murali, V. (2014) 'A Novel Approach for Classification Of Abnormalities In Digitized Mammograms', *Indian Academy of Sciences*', vol. 39, pp. 1141-1150.

(Sachdeva et al. 2013) Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N. and Ahuja, C. K. (2013) 'Segmentation, Feature Extraction, and Multiclass Brain Tumor Classification', *J Digit Imaging*, vol. 26, pp. 1141–1150.

(Sharma et al., 2010) Sharma, G. N., Dave, R., Sanadya, J., Sharma, P. and Sharma, K. K. (2010) 'Various types and management of breast cancer : an overview', *Adv Pharm Technol Res*, vol. 1, pp. 109-126.

(Shastri et al., 2018) Shastri A.A., Tamrakar D. and Ahuja, K. (2018) 'Density-Wise Two Stage Mammogram Classification using Texture Exploiting Descriptors', *Expert Systems with Applications*, vol. 99, pp.71-82.

(Sheba and Raj, 2018) Sheba, K. U. and Raj, S. G. (2018) 'An approach for automatic lesion dandection in mammograms', *Cogent Engineering*, vol. 5, doi : 10.1080/23311916.2018.1444320.

(Shi et al., 2013) Shi, W., Caballero, J., Ledig, C., Zhuang, X. and Rueckert, D. (2013) 'Cardiac image super-resolution with global correspondence using multi-atlas patch match', *Med. Image Comput. Comput. Assist. Interv*, vol. 16, pp. 9-16.

(Shi et al., 2015) Shi, F., Cheng, J., Wang, L., Yap, P. and Shen, D., (2015) 'LRTV : Mr image super-resolution with low-rank and total variation regularizations', *IEEE transactions on medical imaging*, vol 34, pp. 2459-2466.

(Singh and Singh, 2016) Singh, D. and Singh, M. (2016) 'Classification of Mammograms using Support Vector Machine', *International Journal of Signal Processing*', *Image Processing and Pattern Recognition*, vol. 9, pp. 259-268.

(Sklansky, 1978) Sklansky, J. (1978) 'Image segmentation and feature extraction', *IEEE transaction on systems, Man and Cybernetics*, vol. 8, pp. 237-247

- (Srinivasan and Shobha, 2008) Srinivasan, G. N. and Shobha, G. (2008) 'Statistical texture analysis', Proceedings of world academy of science, engineering and technology, vol. 36, pp. 1264-1269.
- (Stone, 1974) Stone M. (1974) 'Cross-validatory choice and assessment of statistical predictions', Journal of the Royal Statistical Society, Series B, vol. 36, pp. 111-147.
- (Subbhuraam et al., 2014) Subbhuraam, V. S., Eddie, Y. K. N., Rajendra, U. A. and Oliver, F. (2014) 'Breast imaging : A survey', World journal of clinical oncology, vol. 2, pp. 171-178.
- (Suresha, 2017) Suresha M., Harisha and Naik T. A. (2017) 'Survey on Image Analysis based on Texture', International Journals of Advanced Research in Computer Science and Software Engineering, vol. 7, pp. 686-695,
- (Suzuki et al., 2016) Suzuki, S., Zhang, X., Homma, N., Ichiji, K., Sugita, N., Kawasumi, Y., Ishibashi, T. and Yoshizawa, M. (2016) 'Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis', Proceedings of the SICE Annual Conference, Tsukuba, Japan, pp. 1382-1386.
- (Tangaro et al., 2008) Tangaro, S., Bellotti, R., Carlo, F., Gargano, G., Lattanzio, E., Monno, P., Massafra, R., Delogu, P., Fantacci, M., Retico, A., Bazzocchi, M., Bagnasco, S., Cerello, Po., Cheran, S., Torres, E., Zanon, E., Lauria, A., Sodano, A., Cascio, D. and Cataldo, R. (2008). 'MAGIC-5 : An Italian mammographic database of digitised images for research', La Radiologia medica, vol.113, pp. 477-485, doi : 10.1007/s11547-008-0282-5.
- (Temizel and Vlachos, 2006) Temizel, A. and Vlachos, T. (2006) 'Wavelet domain image resolution enhancement', IEEE Proc. Vis. Image Signal Process, vol. 153, pp. 25-30.
- (Tfaily et al., 2020) Tfaily, M. A., Nassar, F., Sellam, L. S., Amir-Tidadini, Z. C., Asselah, F., Bourouba. M. and Rihab, N. (2020) 'miRNA expression in advanced Algerian breast cancer tissues', PLOS ONE, vol. 15, doi : 10.1371/journal.pone.0227928.
- (Thibault, 2009) Thibault, G. (2009) 'Indices de formes et de textures : de la 2D vers la 3D Application au classement de noyaux de cellules', thèse de doctorat, université Aix-marseille.
- (Thornton et al., 2006) Thornton, M.W., Atkinson, P.M. and Holland, D.A. (2006) 'Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping', International Journal of Remote Sensing, vol. 27, pp. 473-491, doi : 10.1080/01431160500207088.

- (Trinh et al., 2014) Trinh, D. H, Luong, M. Dibos, F., and al. (2014) ‘Novel example-based method for super-resolution and denoising of medical images’, *IEEE Transactions on Image Processing*, vol. 23, pp. 1882-1895, doi : 10.1109/TIP.2014.2308422.
- (Umehara et al., 2017) Umehara, K., Ota, J. and Ishida, T. (2017) ‘Super-Resolution Imaging of Mammograms Based on the Super-Resolution Convolutional Neural Network’, *Open Journal of Medical Imaging*, vol. 7, pp. 180-195, doi : 10.4236/ojmi.2017.74018.
- (Unser, 1986) Unser, M. (1986) ‘Sum and difference histogram for texture classification’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.8, pp.118-125.
- (Uyun et al., 2013) Uyun, S., Hartati, S., Harjoko, A. and Subanar, S. S. (2013) ‘Selection Mammogram Texture Descriptors Based on Statistics Properties Backpropagation Structure’, *International Journal of Computer Science and Information Security*, vol. 11, pp. 1-5.
- (Vaithyanathan et al., 2013) Vaithyanathan, V., Rajeswari, Tajane, K. and Pitale, R. (2013) ‘Comparison of different classification techniques using different datasets’, *International Journal of Advances in Engineering, Technology*, vol. 6, pp. 764-768.
- (Valarmathie et al., 2016) Valarmathie, P., Sivakrithika, V., Dinakaran, K. (2016) ‘Classification of mammogram masses using selected texture, shape and margin features with multilayer perceptron classifier’, *Biomedical Research, Special issue: Computational Life Science and Smarter Technological Advancement*, pp. S310-S314.
- (Vincent et al., 2010) Vincent, P., Larochele, H., Lajoie, I., Bengio, Y. and Manzagol P.A. (2010) ‘Stacked denoising au to encoders: Learning useful representations in a deep network with a local denoising criterion’, *Journal of Machine Learning Research*, vol. 11, pp. 3371-3408.
- (Wallach and al., 2012) Wallach, D., Lamare, F., Kontaxakis, G., Visvikis, D. (2012) ‘Super-resolution in respiratory synchronized positron emission tomography’, *IEEE transactions on medical imaging*, vol.31, pp. 438-448, doi : 10.1109/TMI.2011.2171358.
- (Wallach, 2011) Wallach, D. (2011) ‘Compensation du mouvement respiratoire en TEP/TDM à l’aide de la super résolution’, thèse de doctorat, université de Bretagne occidentale-Brest.
- (Witten and Frank, 2002) Witten, H., Frank, E. (2002) ‘Data Mining : Practical Machine Learning Tools and Techniques with Java Implementation’, *SIGMOD Rec*, vol. 31, pp. 76-77, doi : 10.1145/507338.507355.

- (Wolberg and Mangasarian, 1990) Wolberg, W. H. and Mangasarian, O. L. (1990) 'Multi surface method of pattern separation for medical diagnosis applied to breast cytology', *Proceedings of the national academy of sciences*, vol. 87, pp. 9193-9196.
- (Xie et al., 2015) Xie, W., Li, Y. and Ma, Y. (2015) 'Breast mass classification in digital mammography based on Extreme learning machine', *Neurocomputing*, doi : 10.1016/j.neucom.2015.08.048
- (Xu et al., 2004) Xu, D. H., Kurani, A., Furst, J. D. and Raicu, D. S. (2004) 'Run-Length Encoding for Volumetric Texture', *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing*, Marbella.
- (Yadollahpour and Shoghi, 2015) Yadollahpour, A. and Shoghi, H. (2015) 'Early Breast Cancer dandection using Mammogram Images : A Review of Image Processing Techniques', *Biosciences Biotechnology Research Asia*, vol. 12, pp. 225-234.
- (Yahiaoui, 2019) Yahiaoui, M. L. (2019) 'les techniques de super-resolution appliquées en imagerie astronomique et nucléaire', thèse de doctorat, université Hadj Lakhdar - BATNA 1.
- (Yue et al., 2016) Yue, L., Shen, H., Li, J., Yuan., Q., Zhang, H. and Zhang, L. (2016) 'Image super-resolution : The techniques, applications, and future.' *Signal Process*, vol. 128, pp. 389-408.
- (Zari, 2019) Zari, A., (2019) 'Uncovering the therapeutic potential of uncaria tomentosa using B16-BL6 mouse melanoma and 4T1 mouse breast cancer cells', thèse de doctorat, université Laurentienne Sudbury, Ontario, Canada.
- (Zhang et al., 2012) Zhang, W., Wang, X., Zhao, D. and Tang, X. (2012) 'Graph Degree Linkage : Agglomerative Clustering on a Directed Graph', In : Fitzgibbon A., Lazebnik S., Perona P., Sato Y. and Schmid C. (eds) *Computer Vision–ECCV 2012, ECCV 2012, Lecture Notes in Computer Science*, vol. 7572, Springer, Berlin, Heidelberg, doi : 10.1007/978-3-642-33718-5_31.
- (Zhang, 2014) Zhang, X., S. (2014) 'A New Approach for Clustered MCs Classification with Sparse Features Learning and TWSVM', *The Scientific World Journal*, doi : 10.1155/2014/970287.
- (Zheng et al., 2007) Zheng, C., Sun, D. W. and Zheng, L. (2007) 'A new region-primitive method for classification of colourmeat image texture based on size, orientation and contrast',

Meat Science, vol. 76, pp. 620-627.

(Zheng et al., 2010) Zheng, J., Fuentes, O. and Leung, M. (2010) 'Super-resolution of mammograms', 2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, Montreal, QC, pp. 1-7, doi : 10.1109/CIBCB.2010.5510384.

Références Web

(Web 01) <https://www.teresewinslow.com>

(Web 02) <https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics>

(Web 03) <https://cdn.imagerieparis13.fr/wp-content/uploads/2018/11/irm-mammaire-irm-paris-imagerie-paris-13-radiologie-irm-scanner-radiographie-echographie-doppler-osteodensitometrie-senologie-infiltration-paris-13-3.jpg>

(Web 04) <https://www.docteur-benchimol.com/la-tomosynthese-mammaire-ou-mammographie-3d.html>

(Web 05) <https://densebreast-info.org/breast-mammography-tomosynthesis.aspx>

(Web 06) <https://www.bci.org.au/breast-cancer-information/fact-sheets/hormonal-therapies-breast-cancer/>

(Web 07) <http://www.eng.usf.edu/cvprg/mammography/database.html>

(Web 8) <https://www.mammoimage.org/databases/>

(Web 09) <http://www.wiau.man.ac.uk/services/mias/miasmini.html>

(Web 10) <https://dept-info.labri.fr/enseignement/programmation2/intro-eclipse/#introtitre>

(Web 11) <https://www.grafikart.fr/tutoriels/netbeans>

(Web 12) <https://www.jmdoudoux.fr/java/dej/chap-presentation.htm>

(Web 13) <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>

Liste des publications

Publications Internationales

- Sawsen Boudraa, Melouah Ahlem, Merouani Hayet Farida, “Improving mass discrimination in mammogram-CAD system using texture information and super-resolution reconstruction”, *Evolving Systems*, 2020, vol. 11, N° 4, pp. 697–706, doi : [//doi.org/10.1007/s12530-019-09322-4](https://doi.org/10.1007/s12530-019-09322-4).

Communications Internationales

- Sawsen Boudraa, Melouah Ahlem, Merouani Hayet Farida, “Une nouvelle combinaison de caractéristiques pour la classification de masses mammaires”, *Actes des 24 èmes Rencontres de la Société Francophone de Classification*, pp. 41-44, 28-30 juin, 2017, Lyon, France.
- Sawsen Boudraa, Melouah Ahlem, Merouani Hayet Farida, “Statistics approach for breast mass classification”, *3rd International Conference on Electrical Engineering and Control Applications*, 21-23 Novembre, 2017, Constantine, Algérie.
- Sawsen Boudraa, Melouah Ahlem, Merouani Hayet Farida, “Classification of mammography image using a new combination of statistics features”, *International Conference on Automatic control, Telecommunications and Signals*, 11-12 Décembre, 2017, Annaba, Algérie.
- Sawsen Boudraa, Melouah Ahlem, Merouani Hayet Farida, “Deep texture representation for breast mass classification”, *4th IEEE International Conference on Signal, Image, Vision and their Applications*, 26-27 Novembre, 2018, Guelma, Algérie, doi : [10.1109/SIVA.2018.8661052](https://doi.org/10.1109/SIVA.2018.8661052).