

الجمهورية الجزائرية الديمقراطية الشعبية
Democratic And Popular Republic of Algeria
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education And Scientific Research

Ziane Achour University of Djelfa
Faculty of Exact and Computer Sciences
Department of Mathematics and Computer Science



جامعة زيان عاشور بالجلفة
كلية العلوم الدقيقة والإعلام الآلي
قسم الرياضيات والإعلام الآلي

Doctoral thesis

Domain: Mathematics & Computer Sciences

Field: Computer Science

Option: Mathematical Optimization for Signal Processing and Communication Networks

By:

Mr. Nasser edinne BENHASSINE

Contributions to discriminant feature extraction methods for classification of medical images

Defended on: July 25th, 2019

Jury Members:

Mr. Sami SAADI	Professor	Univ of Djelfa	President
Mr. Djalil BOUDJEHEM	Professor	Univ of Guelma	Supervisor
Mr. Abdelnour BOUKAACHE	MC-A	Univ of Guelma	Co-Supervisor
Mr. Farouk CHELLALI	MC-A	Univ of Djelfa	Examiner
Mr. Ahmed GAZEL	MC-A	Univ of Djelfa	Examiner
Mr. Djamel ABED	MC-A	Univ of Guelma	Examiner

Academic Year: 2018-2019

الجمهورية الجزائرية الديمقراطية الشعبية
Democratic and popular republic of Algeria

وزارة التعليم العالي والبحث العلمي
Ministry of higher education and scientific research

Université Ziane Achour de Djelfa
Faculté des sciences Exactes & Informatique
Département Mathématique & Informatique



جامعة زيان عاشور بالجلفة
كلية العلوم الدقيقة والإعلام الآلي
قسم الرياضيات والإعلام الآلي

THÈSE DE DOCTORAT

Domain: Mathématiques & Informatique

Filière: Informatique

Spécialité: optimisation mathématique pour le traitement du signal et les réseaux de communication

Présenter Par:

Mr. Nasser edinne BENHASSINE

Contributions aux méthodes d'extraction de caractéristiques discriminantes pour la classification des images médicales

Soutenu le: 25 juillet 2019

Membres du jury

Mr. Slami SAADI	Professeur	univ de Djelfa	Président
Mr. Djalil BOUDJEHEM	Professeur	univ de Guelma	Rapporteur
Mr. Abdelnour BOUKAACHE	MC-A	univ de Guelma	Co-rapporteur
Mr. Farouk CHELLALI	MC-A	univ de Djelfa	Examineur
Mr. Ahmed GAZEL	MC-A	univ de Djelfa	Examineur
Mr. Djamel ABED	MC-A	univ de Guelma	Examineur

Année académique: 2018/2019

Dédicaces

Je dédie cette thèse

À ma mère et à mon père, que Dieu ait pitié de lui, qui m'a toujours aimé inconditionnellement et qui m'a appris à travailler dur pour atteindre les objectifs auxquels j'aspire.

À ma femme et à mes enfants,

Djihad, Nadir, Faysal, Amani, batta et le petit Naceredinne

À mes chers frères et sœurs et leurs familles

à tous mes amis

Et tous ceux qui ont contribué un jour à l'éducation et à la formation.

Remerciements

Tout d'abord nous adressons notre gratitude sincère à notre Dieu tout puissant de m'avoir donné le courage, la force et la patience.

Je remercie mes encadreurs Pr. Boudjehem Djalil et Dr. Boukaache Abdenoure pour avoir dirigés ma thèse avec beaucoup d'efforts et de patience et ses précieux conseils. Ils furent tout le temps attentifs et disponible malgré ses nombreuses charges.

j'exprime aussi mes vifs remerciements aux membres du jury d'avoir accepté d'évaluer ce travail.

De plus, j'aimerais remercier tous les membres du comité de formation des doctorants de nous avoir donné l'occasion d'exploiter nos capacités.

Je tiens à remercier toute personne ayant contribué de près ou de loin à la réalisation de ce Travail.

Benhassine Nasser eddine

ملخص

في هذه الأطروحة تم اقتراح وإنجاز نظام تشخيصي جديد بمساعدة الكمبيوتر لتصنيف أورام الثدي في صور الماموغرام (صور الثدي بالأشعة السينية) على أنها خبيثة أو حميدة. يتكون النظام المقترح من ثلاث خطوات أساسية. أولاً: استخدام مجموعة من عمليات المعالجة الأولية لاستخراج المنطقة المهمة (ROI) المراد دراستها والتي تحتوي فقط على الثدي. تم حذف عضلة الصدر باستخدام طريقة جديدة تدمج بين خوارزمية تجزئة المنطقة (SRG) وخوارزمية تحديد المنطقة المثلثية التي تحتوي فقط على العضلة. ثانياً للحصول على الصفات المميزة لأورام الثدي يتم تحويل جميع المنطقة المهمة (ROI) التي تم الحصول عليها باستعمال تحويل جيب التمام المنفصل (DCT) أو تحويل الموجات المنفصلة (DWT)، ثم يتم اختيار بعض المعاملات المحولة باستخدام طريقة تحليل قوة التمييز (DPA). أخيراً، يتم استخدام الخصائص المحددة بواسطة المصنفات الآتية: شعاع الدعم الآلي (SVM)، الشبكات العصبية الاصطناعية (ANN)، مصنف بايز (Bayesian Naïve)، ومصنف ك-أقرب الجيران (KNN) لتصنيف صور الماموغرام لقاعدة بيانات MIAS إلى أورام حميدة أو خبيثة.

الكلمات المفتاحية: التشخيص بمساعدة الكمبيوتر، التصنيف، تصوير الثدي بالأشعة السينية، DCT، DPA، ANN، SVM

Résumé

Dans cette thèse, un nouveau système de diagnostic assisté par ordinateur est proposé pour classer les tumeurs du sein dans les mammographies comme malignes ou bénignes. Le système proposé comprend trois étapes. Tout d'abord, un ensemble d'opérations de prétraitement sont utilisés pour extraire la région d'intérêt (ROI) contenant le sein. L'élimination du muscle pectoral est effectuée à l'aide de l'algorithme de segmentation SRG (*Seeded Region Growing*) appliqué dans une région triangulaire pré-localisée. Ensuite, toutes les ROI obtenues sont transformées par la transformée en cosinus discrète (DCT) ou la transformée en ondelettes discrète (DWT), puis certains coefficients transformés sont sélectionnés à l'aide d'une méthode d'analyse de la puissance de discrimination (DPA) de ces coefficients. Enfin, les caractéristiques sélectionnées sont utilisées par les classificateurs SVM (*Support Vector Machine*), ANN (*Artificial Neural Networks*), NB (*Naïve Bayésien*) et KNN (*K-Nearest Neighbors*) pour classer les mammographies de la base de données MIAS en tumeurs bénignes ou malignes.

Mots-clefs: diagnostic assisté par ordinateur, mammographie, DCT, DPA, ANN, SVM.

Abstract

In the thesis we proposed and implemented a computer aided diagnosis (CAD) system to assist radiologists in the classification of mammogram images. The CAD system is composed of three main steps. The first step is image pre-processing and segmentation with the seeded region growing (SRG) algorithm applied on a localized triangular region to remove only the muscle. In the second step of the CAD system, we proposed a novel features extraction method, which consists of three stages. In the first, the discrete cosine transform (DCT) is applied on all obtained regions of interest (ROI) and then only the upper left corner (ULC) of DCT coefficients is retained. Secondly, we have applied the energy probability (EP) to the ULCs that is used as a criterion for selecting discriminant information. At the last stage, a new Most Discriminative power coefficient algorithm (MDPEP) has been proposed to select the most significant features. In the final step of the CAD, the Support Vector Machines (SVM), Naive Bayes (NB) and Artificial Neural Network (ANN) classifiers are used to make an effective classification. The evaluation of the proposed algorithm on the mini-MIAS database shows its efficiency over other recently proposed CAD systems in the literature

Key words: computer aided diagnosis, classification, Mammogram, DCT, DPA, ANN, SVM.

Tableau des matières

Dédicaces	I
Remerciements	II
Résumé	III
Tableau des matières.....	IV
Liste des Tableaux.....	VII
Liste des Figures	IX
Index des symboles	X
Introduction Générale	1
CHAPITRE I Images médicales	4
I.1 Introduction	4
I.2 Imagerie médicale	5
I.2.1 Histoire de l'imagerie médicale	5
I.2.2 Les intérêts de l'imagerie médicale.....	6
I.2.3 Les grands types d'imagerie médicales	6
I.2.3.1 La Radiographie	6
I.2.3.2 Scanner	8
I.2.3.3 Échographie.....	8
I.2.3.5 La médecine nucléaire.....	10
I.3 Mammographie et cancer du sein	12
I.3.1 Anatomie du sein	12
I.3.2 Le cancer du sein.....	13
I.3.3 La mammographie	14
I.3.3.1 Principe	14
I.3.3.2 Limites de la mammographie.....	15
I.3.3.3 Le mammographe	15

I.3.3.3 Relation entre les images de mammographie et l'anatomie du sein	16
I.3.3.4 Les incidences en mammographie.....	17
I.3.3.5 Les examens mammographiques	18
I.4 Les différentes lésions mammaires	18
I.4.1 Les calcifications	19
I.4.1.1 Micro calcification.....	19
I.4.1.2 Les macrocalcifications.....	20
I.4.2 Les masses	20
I.4.2.1 La forme	21
I.4.2.2 Le contour	21
I.4.3 Densité mammaire (DM)	21
I.4.4 La classification des anomalies mammographiques.....	23
I.5 Diagnostic Assisté par Ordinateur (CAD).....	24
I.6 Conclusion.....	24
CHAPITRE II Système d'aide au diagnostic	26
II.1 Introduction.....	26
II.2 Définition	27
II.3. Etat de l'art des CAD pour le cancer du sein.....	27
II.4 Les étapes de développement d'un système d'aide au diagnostic	29
II.4.1 Prétraitement des images.....	30
II.4.1.1 La réduction du bruit	30
II.4.1.2 Elimination des artefacts.....	34
II.4.2 Segmentation	36
II.4.2 .1Définition.....	36
II.4.2 .2 Les approches de segmentation.....	36
II.4.2.3 Segmentation du muscle pectoral.....	41
II.4.3 Extraction de caractéristiques	42

II.4.4 Sélection de caractéristiques.....	44
II.4.5 Classification des anomalies.....	44
II.5 Conclusion.....	45
Dans ce chapitre nous avons donné un aperçue générale sur les systèmes d'aide aux décisions et leurs différentes parties, ainsi que les techniques utilisées dans le domaine d'imagerie médicale et plus précisément dans les images mammographie.....	45
nous avons cité un état de l'art sur les efforts considérables de nombreux chercheurs pour aider les radiologues en mettant au point des outils d'interprétation assistée par ordinateur pour la mammographie.....	45
CHAPITRE III La classification des images médicales	47
III.1 Introduction	47
III.2 Définition.....	48
III.3 Les Méthodes de Classification.....	48
III.3.1 Méthodes de Classification supervisées.....	49
III.3.1.1 K-Plus Proches Voisins (KNN)	50
III.3.1.2 Séparateur à Vaste Marge (SVM).....	53
III.3.1.3 Naïve bayésienne	55
III.3.1.4 Réseau de neuronne.....	56
III.3.2 Méthodes de Classification Non Supervisées.....	58
III.4 Conclusion.....	59
CHAPITRE IV Implémentation et résultats.....	61
IV.1 Introduction	61
IV.2 Description des bases de données (cancer du sein).....	62
IV.2.1 La base de donnée (MIAS).....	62
IV.2.2 Informations détaillées de la base de données MIAS.....	62
IV.2.3 Echantillons de la base (MIAS).....	64
IV.3 le système d'aide au diagnostic (CAD) proposé	66
IV.3 .1 Létape de prétraitement.....	66
IV.3.2 Extraction et sélection de caractéristiques.....	71

IV.3.2.1 Première partie (domaine fréquentiel DCT)	73
IV.3.2.2 deuxième Partie (domaine ondelettes DWT).....	76
IV.3.3 Classification.....	78
IV.4 Résultats et discussions	80
IV.4.1 Résultats du première partie	81
IV.4.2 Résultat du deuxième Partie (domaine DWT).....	85
IV.6 Conclusion.....	89
Conclusion générale et perspectives	90
Liste des productions scientifiques.....	92
Bibliographie.....	93

Liste des Tableaux

Tableau I.1 classification BIRADS.	22
Tableau II.1 Avantages et inconvénients de méthode par croissance de régions.	39
Tableau II.2 Comparaison des techniques de segmentation.	41
Tableau III.1 Analogie entre les neurones biologiques et artificiels.	57
Tableau III.2 Exemple de fonctions d'activation (comparées au seuil θ).	57
Tableau IV.1 Partie de liste répertorie des images de la base MIAS.	62
Tableau IV.2 Les résultats des performances de classification obtenus (ANN, SVM, NB, KNN)	81
Tableau IV.3 performances pour différents tailles (ULC) et différents nombres de caractéristique.	82
Tableau IV.4 Performances de classification de KNN	84
Tableau IV.5 Performances de classification en fonction du nombre d'images (malins, bénins)	85
Tableau IV.6 Les résultats des performances de classification obtenus.	86
Tableau IV.7 Les résultats des performances de classification pour SVM.	86
Tableau IV.8 Les résultats des performances de classification pour ANN.	86
Tableau IV.9 Les résultats des performances de classification pour NB.	87
Tableau IV.10 Comparaison de Résultats avec les méthodes existantes.	88

Tableau des figures

Figure I.1 Equipements radiographiques.....	7
Figure I.2 Exemple images (radiographie).	7
Figure I.3 Image d'une échographie obstétricale.....	9
Figure I.4 Exemple d'un échographe.....	9
Figure I.5. Exemple images (IRM).....	10
Figure I.6 Gamma caméra couplée à un système scanner.....	11
Figure I.7. Anatomie du sein.	13
Figure I.8 Résultat d'une mammographie.....	15
Figure I.10 Illustrations des incidences principales en mammographie	17
Figure I.11 Les différents types de microcalcifications	20
Figure I.12 Les différentes formes possibles d'une masse.....	21
Figure I.13 Les différents types de contour d'une masse:.....	21
Figure I.14 Différents types de densité mammaire.....	22
Figure II.1 système d'aide à la décision.....	27
Figure II.2 Diagramme général d'un système CAD.....	29
Figure II.3 Système de prétraitement.	30
Figure II.4 Le filtre Moyen.....	31
Figure II.5 filtre Médian (la valeur centrale 80 est remplacée par 11).....	33
Figure II.6 Décomposition d'ondelettes à différents niveaux (DWT).....	34
Figure II.7 Exemple d'une image mammographies contient radiopaque artefacts:.....	35
Figure II.8 Exemples d'étiquettes du film mammographique:	35
Figure II.9 Exemples de marqueurs opaques sur une mammographie.....	35
Figure II.10 les techniques de segmentation d'image.	37
Figure II.11 Exemple seuillage global	38
Figure II.12 partitionnement d'image par division en utilisant "quad-tree".....	39
Figure II.13 regroupement des régions similaires suivant un critère de fusion.....	40
Figure II.14 Vue générale de descripteur d'image.....	43
Figure III.1 Différents techniques de classification.....	49
Figure III.2 illustration simple par un algorithme KNN (K=5).	51
Figure III.3 Représentation de l'hyperplan optimal pour des données séparables.....	54
Figure III.4 Structure d'un neurone artificiel.....	56

Figure III.5 Architecture d'un réseau de neurones (avec deux couches cachées).....	58
Figure IV.1 Caractère du tissu de fond.....	65
Figure IV.2 Diagramme du système proposé.....	66
Figure IV.3 Réduction du bruit avec le filtre médian.....	67
Figure IV.4 Suppression des artefacts et séparation de fond.....	68
Figure IV.5 Représentation de la localisation triangulaire du muscle pectoral.....	69
Figure IV.6 Les étapes de suppression du muscle pectoral.	70
Figure IV.7 Suppression du muscle pectoral.....	70
Figure IV.8 Création du masque binaire.....	75
Figure IV.9 Diagramme du système proposé dans le domaine en ondelette.....	76
Figure IV.10 décomposition en 2 niveaux avec <code>dwting()</code> [95].....	77
Figure IV.11 la précision de classification par rapport à la taille ULC (avec 100 caractères)...	83
Figure IV.12 précision de la classification en fonction du nombre de caractères.....	83
Figure IV.13 performance de classification du ANN[94].....	84

Index des symboles

ACR: American College of Radiology.
AMDI: Indexed Atlas of Digital Mammograms.
ANN: Artificial Neural Network.
BI-RADS: Breast Imaging-Reporting And Data System.
CA: Approximation Coefficients.
CAD: Computer Aided Diagnosis system.
CADe: Computer Aided Detection system.
CC: Cranio Caudale.
CD: Détails Coefficients.
DCT: Discrete Cosine Transform.
DDSM: Digital Database for Screening Mammography.
DM: Densité mammaire.
DPA: Discriminant Power Analysis.
DWT: Discrete Wavelet Transform.
EP: Energy Probabilty
GLCM: Grey Level Cooccurrence Matrix.
GLRLM: Grey Level Run Length matrix.
IDCT: Inverse Discrete Cosine Transform .
IDWT: Inverse Discrete Wavelet Transform.
IRM: Imagerie Resonance Magnetic.
KNN: K-Nearest Neighbors.
LTRPM : Localisation of the Triangular Region of Pectoral Muscle.
MDP : Most Discrimination Power matrix.
MDPEP: Most Discriminative Power of coefficients Energy Probility algorithm.
MIAS: Mammographic Image Analysis Society.
MLO: Medio Lateral Oblic.
MLP: Multi Layer Perceptron.
NB: Naïve Bayes
OMS: Organisation Mondiale de Santé
ROC: Receiver Operating Characteristic.
ROI: Region Of Interest.
SOM: Self-Organizing Map (réseau de neurones de kohonen).
SRG: Seeded Region Growing.
SVM: Support vectors machine.
ULC: Upper Left Corner.
WDBC: Wisconsin Diagnosis Breast Cancer.

Introduction Générale

Le cancer du sein reste le type le plus meurtrier, il est devenu un grave problème de santé publique dans de nombreux pays. On estime que plus de 571 000 femmes dans le monde sont décédées du cancer du sein en 2015 [79]. Actuellement, la mammographie est utilisée pour le diagnostic précoce du cancer du sein. D'autre part, la croissance rapide des mammographies a créé un besoin pour développer des systèmes de diagnostic assisté par ordinateur (CAD) aidant les radiologues à la détection précoce en classifiant une tumeur du sein comme maligne ou bénigne [80] - [81]. En imagerie médicale, de nombreuses études sur le cancer du sein ont montrés qu'une détection précoce et un traitement approprié de la maladie pouvaient réduire le taux de mortalité. Actuellement, la mammographie est l'outil le plus efficace dans la détection précoce du cancer du sein. En outre, la mise en œuvre des systèmes CAD contient divers étapes tels que l'amélioration des images de mammographies, l'identification des régions d'intérêt après segmentation des images, l'extraction et la sélection des de caractéristiques et finalement la classification des mammographies.

L'extraction et la sélection des caractéristiques est un processus très important dans un système CAD, elles influent directement sur les performances du système dans la classification et la reconnaissance des formes. Ils existent différents méthodes d'extraction de caractéristiques à partir de l'image. Dans la littérature, de nombreuses techniques ont été développées pour la détection et le diagnostic du cancer du sein [17, 18, 19, 20, 21].

Dans cette thèse, un nouveau système d'aide à la diagnostic assisté par ordinateur (CAD) a été proposé et mis en œuvre pour classer les tumeurs du sein dans les mammographies comme malignes ou bénignes. Premièrement, pour l'identification de la région d'intérêt, nous proposons l'application de la technique de croissance de régions (SRG) dans une région triangulaire présélectionnée pour éliminer le muscle pectoral.

Après, nous appliquons une transformation des régions d'intérêt (ROI) en utilisant des transformations très connues et utilisées dans le domaine de traitement des images, notamment la transformée en cosinus discrète (DCT) et la transformée en ondelettes discrète (DWT). Ensuite, nous proposons la sélection des caractéristiques les plus significatives à l'aide d'une analyse discriminante (DPA) pour générer un masque identifiant les positions des

caractéristiques les plus importantes dans le domaine transformé. Enfin, les classifieurs SVM, ANN, KNN et NB sont utilisés pour classer les images de mammographies de la base de données MIAS [75].

La thèse est organisée en quatre chapitres qui nous permettront de présenter les différents aspects de notre travail.

Dans le premier chapitre nous introduisons des notions générales sur les modalités d'acquisition des images médicales, en donnant l'anatomie du sein et la définition du cancer du sein.

Le deuxième chapitre sera dédié à la présentation des systèmes automatique d'aide à la détection/diagnostic, ces caractéristiques, ces étapes par lesquelles passe un système de classification de cancer du sein, ainsi que les approches utilisées pour sa réalisation.

Le troisième chapitre donne un aperçu général sur la classification d'images et les différentes approches utilisées.

Le dernier chapitre contient le fruit de notre travail, il présente l'implémentation de notre système (CAD) proposé et les résultats obtenus sous forme de Tableaux, de graphiques, de courbes, de comparaisons et de toutes les analyses et discussions nécessaires.

CHAPITRE I

Images médicales

CHAPITRE I Images médicales

I.1 Introduction

Le développement rapide des technologies d'imagerie médicale est entrain de révolutionner la médecine. L'imagerie médicale permet de représenter sur pellicule différentes parties de l'organisme, et de voir l'intérieur du corps humain, la structure ou la fonction des organes, sans procéder à une intervention directe sur ceux-ci. L'imagerie médicale est utilisée afin d'étayer un diagnostic ou de proposer un traitement pour de nombreuses pathologies, notamment en cancérologie. En 1895 avec l'utilisation des rayons X, c'est le début de la radiographie, et la première application d'imagerie médicale.

Les techniques d'images apparaissent au long du XXème siècle suite à de nombreuses découvertes en physique. Elles s'améliorent et se diversifient avec l'arrivée de la scintigraphie, du scanner, de l'échographie puis de l'IRM.

L'imagerie médicale peut être utilisée par exemple dans le dépistage systématique des cancers du sein par mammographie, ou pour confirmer ou infirmer un diagnostic supposé.

L'imagerie médicale est essentielle pour:

- La recherche médicale.
- L'étude des maladies.
- La création et l'administration de traitements.

Dans ce chapitre nous nous commençons par citer les techniques d'imagerie médicale qui utilisent la radioactivité, les rayons X, les ultrasons ou encore la résonance magnétique, ainsi qui forment ce que l'on appelle la médecine nucléaire (scintigraphie, tomographie par émission de photons et tomographie par émission de positons), Ensuite, nous présentons brièvement l'anatomie du sein chez la femme, et la notion du cancer du sein et les caractéristiques des lésions mammaires. Ensuite, nous abordons les outils d'imagerie médicale permettant le dépistage et le diagnostic de ce type de cancer notamment la mammographie, Puis, nous citons la relation qui existe entre la composition des structures anatomiques du sein et leur apparence sur une image mammographique. Finalement nous indiquons le besoin d'automatiser l'interprétation des mammographies par les radiologues.

I.2 Imagerie médicale

L'imagerie médicale est un ensemble des moyens d'acquisition d'images, et de techniques consistant à mettre en image différentes régions du corps humain. Ces techniques basés sur plusieurs phénomènes physiques tels que la résonance magnétique, la radioactivité, l'absorption et atténuation des rayons X, la réflexion d'ondes d'ultrasons, l'effet photoélectrique, etc. Il existe plusieurs types d'imageries médicales qui sont plus ou moins adaptées en fonction des zones à étudier.

On distingue notamment la radiologie, qui utilise les rayons X, l'échographie, qui se sert des ultrasons, l'imagerie par résonance magnétique (IRM), le scanner qui donne également des images en coupe et qui permet de réaliser la représentation d'organes en 3D.

I.2.1 Histoire de l'imagerie médicale

1895. W. Röntgen, physicien allemand, découvre les rayons X.

1896. H. Becquerel, physicien français, découvre que l'uranium émet des rayonnements invisibles (différents des rayons X). Il donne le nom de « radioactivité » à ce phénomène.

1898. P. et M. Curie, physiciens français, isolent le polonium et le radium, deux éléments radioactifs jusqu'alors inconnus, présents dans le minerai d'uranium. Le prix Nobel de physique leur est décerné, ainsi qu'à Henri Becquerel, en 1903.

1913. G. C. de Hevesy, chercheur hongrois, utilise le radium, isotope radioactif naturel, pour en étudier sa distribution dans le corps d'un mammifère, il s'agit du premier traceur.

1928. H. Geiger et W. Müller créent le premier compteur de particules chargées.

1930. E. O. Lawrence, physicien américain, met au point un accélérateur électromagnétique de haute fréquence, procédé qui sera ensuite repris pour le cyclotron.

1935. I. et F. Joliot, physiciens français, reçoivent le prix Nobel pour leur découverte des isotopes radioactifs en 1934. Lors de la conférence de remise du prix, Frédéric Joliot déclare : « La méthode des indicateurs employant des radioéléments synthétiques trouvera probablement des applications pratiques en médecine. » George de Hevesy utilise le phosphore ³² pour montrer que la formation des os est un processus impliquant en permanence des pertes et des remplacements. Ses travaux sur les traceurs seront récompensés en 1943 par un prix Nobel de chimie.

1937. J.-G. Hamilton effectue la première utilisation clinique du sodium radioactif.

1938. S. Hertz utilise de l'iode radioactif (^{131}I) pour l'étude de la physiologie thyroïdienne.

1939. G. C. de Hevesy met au point une méthode de détermination du volume sanguin utilisant des globules rouges marqués au phosphore (^{32}P). J.H. Lawrence utilise ce procédé pour étudier les leucémies.

1942. J.-G. Hamilton effectue les premières applications thérapeutiques de l' ^{131}I et du ^{32}P .

1950. Le premier scanner manuel apparaît, doté d'un compteur Geiger-Müller, puis d'un compteur à scintillations. Cette technique sera améliorée l'année suivante par B. Cassen, qui crée un scanner rectilinéaire.

1957. H.O. Anger invente une caméra à scintillations, gamma-caméra capable de produire des images en corrélation avec le fonctionnement des organes, c'est la scintigraphie.

1963. Godfrey Newbold Hounsfield, ingénieur britannique, met au point des algorithmes mathématiques, qui sont appliqués ensuite par Allan M. Cormack pour les scanners.

1973. Première image d'IRM.

I.2.2 Les intérêts de l'imagerie médicale

L'imagerie médicale a révolutionné la médecine, et devient incontournable dans de nombreuses situations tels que le diagnostic, l'évaluation de la sévérité d'une pathologie et le teste de l'efficacité d'un traitement.

I.2.3 Les grands types d'imagerie médicales

Il existe quatre types d'imagerie médicale qui reposent sur l'utilisation des rayons X, des ultrasons, du champ magnétique ou de la radioactivité

I.2.3.1 La Radiographie

La radiologie est inventée en 1895 par le physicien allemand Wilhelm Conrad Röntgen, auteur des premières radiographies de l'anatomie humaine, son principe repose sur l'utilisation de deux grands équipements : une source de rayons X positionnée devant une région anatomique et un détecteur de rayons X qui est placé de l'autre côté (Figure I.1).

La radiographie est une imagerie par transmission et projection. La source de rayons est à l'extérieur du corps, et chaque point de l'image correspond à une information le long d'une trajectoire linéaire à travers le patient (Figure I.2). Dans ces principaux domaines, la radiographie est très largement utilisée pour le diagnostic de fractures osseuses, de cancer des poumons et de problèmes cardiovasculaires. Elle permet également d'observer des anomalies sur certains organes comme des infections bactériennes ou virales ou encore des tumeurs au niveau des poumons ou des seins (mammographie).



Figure I.1 Equipements radiographiques.



Figure I.2 Exemple images (radiographie).

I.2.3.2 Scanner

Le scanner est découvert par l'ingénieur britannique Godfrey Newbold Hounsfield en 1972, se base sur des techniques informatiques de reconstruction. Il repose également sur l'utilisation des rayons X mais permet d'obtenir des images tridimensionnelles des organes ou des tissus (os, muscles ou vaisseaux) sous forme de coupes. Grâce au scanner on visualise une modification de volume ou une anomalie de structure (infections, hémorragies, tumeur, ganglions, embolie), le scanner mettra en évidence un rétrécissement des artères coronaires en cas de douleurs thoraciques.

I.2.3.3 Échographie

L'échographie, introduite en médecine dans les années 1950, est dérivée du principe du sonar. Les images résultent de l'émission et la réflexion des ultrasons sur les organes pleins de l'abdomen, le cœur et tous les organes non masqués par le squelette (globe oculaire, cerveau chez le nouveau-né) ou par les gaz. L'échographie est utile à l'observation du fonctionnement des organes comme l'appréciation de la vitalité et de la morphologie du fœtus lors de la grossesse. la figure I.3 presente image d'une échographie obstétricale.

Malgré l'échographie a gagné sa renommée dans le suivi des grossesses, elle est très utile en cardiologie, gastro-entérologie (foie, bile, rein) et urologie (appareils génitaux, vessie). Les zones contenant de l'air tel que les poumons, tube digestif et os ne sont pas analysables par cette technique. Les ultrasons ne peuvent être utilisés qu'en tant que modalité de diagnostic limitée à l'imagerie anatomique [1]. Sur la figure I.4 ci-dessous, on presente un exemple d'un échographe.



Figure I.3 Image d'une échographie obstétricale.



Figure I.4 Exemple d'un échographe.

I.2.3.4 Imagerie par résonance magnétique (IRM)

L'imagerie par résonance magnétique (IRM) est apparue au début des années 1980 [2], du phénomène de résonance magnétique nucléaire (RMN) découvert en 1946 par les physiciens américains Edward Mills Purcell et Félix Bloch. Elle permet l'acquisition d'images en coupes, dans tous les plans de l'espace, ainsi que des représentations tridimensionnelles en visualisant différentes structures ; notamment en neuro-imagerie (cerveau, moelle épinière) et les tissus mous (viscères, muscles et tendons) [2]. L' (IRM) permet ainsi de confirmer un diagnostic de sclérose en plaques ou de maladie d'Alzheimer. Sur la figure I.5 ci-dessous, on présente un exemple d'images (IRM).

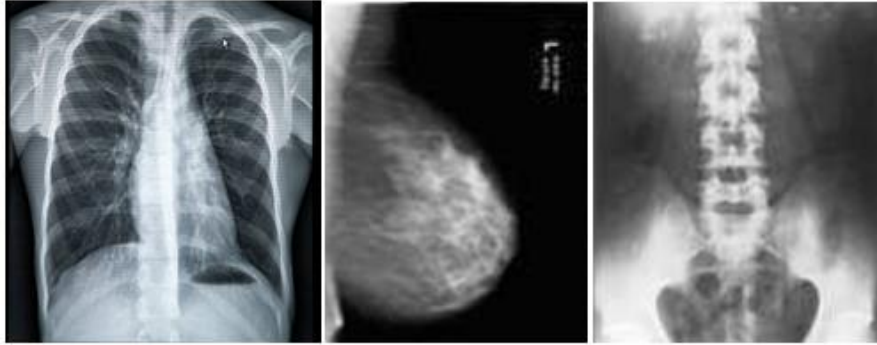


Figure I.5. Exemple images (IRM).

I.2.3.5 La médecine nucléaire

La médecine nucléaire ou isotopique, est issue de la découverte, en 1896, de la radioactivité par Henri Becquerel, dont les travaux furent repris par les physiciens Pierre et Marie Curie. Alors elle est une branche de la radiologie qui désigne soit l'imagerie par résonance magnétique, soit **la scintigraphie**.

Ce sont deux domaines très différents de l'imagerie qui se regroupent pourtant sous le même chapitre. Pour l'imagerie par résonance magnétique, l'origine du nom vient du fait que ce sont des protons d'hydrogènes qui servent à produire l'image. Pour la scintigraphie, on utilise la capacité d'un élément radioactif à se fixer sur des cellules du corps d'où ils émettront des rayonnements qui pourront être détectés.

La médecine nucléaire est avant tout un outil d'imagerie fonctionnelle, elle consiste en l'administration d'un radioélément (isotopes radioactifs) soit pour réaliser une imagerie métabolique à visée diagnostique, soit pour un traitement des pathologies affectant des organes ou des tissus (radiothérapie métabolique). Ces radioéléments émettent différents types de rayonnement qui permet de voir et d'analyser le fonctionnement de divers organes. Ainsi, le médecin peut accéder à une imagerie fonctionnelle et métabolique pour détecter des lésions, surveiller leur évolution, mais aussi assister précisément un éventuel geste chirurgical à pratiquer. L'imagerie de médecine nucléaire est une méthode d'imagerie fonctionnelle qui permet de vérifier et de confirmer la mort ou la vie d'un tissu ou d'un organe. À la différence de toutes les autres modalités d'imagerie tels que l'imagerie de Résonance Magnétique (IRM), Rayons X (RX) ou Ultrasons (US) seront incapables de faire la différence entre les deux états, vivant ou mort [1].

La scintigraphie fait partie des techniques d'imagerie médicale qui repose sur l'injection d'une substance faiblement radioactive dans l'organisme. Cette substance est conçue pour se fixer préférentiellement sur un organe ou une tumeur. La détection des particules qu'elle émet permet d'obtenir une image très fine. Elle permet également de suivre l'activité d'un organe, par exemple le cerveau ou le cœur. Grâce à la scintigraphie on peut repérer des métastases et donc mesurer le niveau de dissémination d'un cancer dans l'organisme.

Pour produire une scintigraphie, on utilise un gamma caméra, qui sert à l'acquisition des images à partir des rayons gamma émis par le matériau radioactif (Figure I.6).



Figure I.6 Gamma caméra couplée à un système scanner.

En utilisant la gamma caméra, on peut obtenir plusieurs types d'images : des images statiques (scintigraphie plane), des images dynamiques qui permettent, par exemple, de voir la circulation du traceur injecté, des balayages du corps entier, des images tomographiques où l'enregistrement se fait autour d'un axe en permettant une reconstruction ultérieure 3D.

Deux techniques de scintigraphie sont actuellement utilisées :

- la tomographie par émission mono photonique (TEMP).
- la tomographie par émission de positons (TEP).

a) La tomographie par émission de positons (TEP)[1]

C'est une méthode d'imagerie médicale pratiquée par les spécialistes en médecine nucléaire qui permet de mesurer en trois dimensions une activité métabolique ou moléculaire d'un organe grâce aux émissions produites par les positons issus d'un produit radioactif injecté au préalable.

La TEP repose sur le principe général de la scintigraphie qui consiste à injecter un traceur dont on connaît le comportement et les propriétés biologiques pour obtenir une image du fonctionnement d'un organe ou la présence d'une cible moléculaire. Ce traceur est marqué par un atome radioactif (carbone, fluor, azote, oxygène...) qui émet des positons dont l'annihilation produit deux photons. C'est la détection en coïncidence de ces photons qui permet la localisation du lieu de leur émission et donc la concentration du traceur en chaque point de l'organe. C'est cette information quantitative que l'on représente sous la forme d'une image faisant apparaître en couleurs les zones de forte concentration du traceur.

b) La tomographie d'émission mono photonique (TEMP) [1]

aussi appelée tomoscintigraphie par émission mono photonique, c'est une technique d'imagerie fonctionnelle qui repose sur le principe de la scintigraphie et qui permet d'effectuer des images ainsi que des reconstructions en trois dimensions d'organes et de leur métabolisme à l'aide d'un ensemble de gamma caméras tournant autour du patient.

I.3 Mammographie et cancer du sein

I.3.1 Anatomie du sein

Le rôle biologique du sein est la production, le stockage et la sécrétion du lait pour l'allaitement du nourrisson, chaque sein est constitué d'une glande mammaire, de tissu conjonctif et de tissu adipeux graisseux qui leur donne sa forme et son volume. La glande mammaire est formée de 15 à 25 lobes, chaque lobe étant composé de vingt à quarante lobules. Ces lobules renferment des unités sécrétrices, les *acini*, qui produisent le lait. Les canaux transportent en suite le lait vers le mamelon. Le sein abrite aussi des vaisseaux et des ganglions qui font partie du système lymphatique, dont le rôle est de combattre les infections.

La figure (Figure I.7) donne un aperçu de la disposition des structures anatomiques importantes du sein.

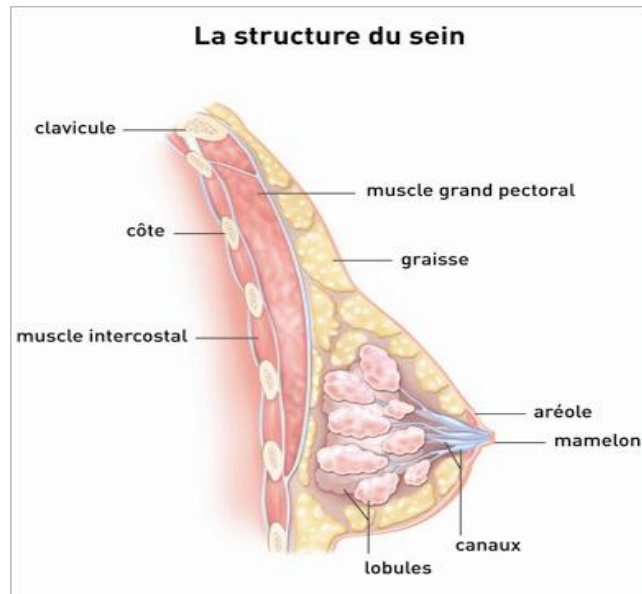


Figure I.7. Anatomie du sein.

I.3.2 Le cancer du sein

Le cancer est une maladie caractérisée par la multiplication et la propagation anarchiques de cellules anormales, dont l'évolution a pour conséquence le décès de la personne concernée [3]. Ce phénomène favorise l'apparition de nouvelles mutations et augmente ainsi le rythme de ces divisions anarchiques. Ces premières cellules cancéreuses forment alors une tumeur maligne s'étant ensuite divisée, formant la tumeur primaire constituée de cellules clonales. Parfois, les cellules cancéreuses envahissent les tissus environnants, ou se détachent de la tumeur d'origine pour migrer vers d'autres régions du corps, c'est ce qu'on appelle des métastases.

Le cancer du sein est le cancer le plus fréquent chez la femme et parmi les premières causes de mortalité dans le monde, il se développe dans les trois quarts des cas chez des femmes de plus de 50 ans, mais il peut apparaître chez des femmes plus jeunes, notamment en cas de prédispositions génétiques. Le cancer du sein représente la principale cause de mortalité chez la femme entre 35 et 65 ans. Le cancer du sein peut aussi concerner l'homme, mais ils ne représentent que 1% des cas de cancer du sein. Il se caractérise par une prolifération anormale de certaines cellules, et naît le plus souvent au niveau de l'épithélium des glandes mammaires. Il peut ensuite s'étendre localement, pour atteindre les ganglions axillaires (au niveau de l'aisselle), via les vaisseaux lymphatiques. La tumeur se développe soit à partir des lobules qui produisent

le lait (carcinome lobulaire), soit, plus fréquemment, à partir des canaux qui conduisent le lait vers le mamelon (carcinome canalaire).

I.3.3 La mammographie

I.3.3.1 Principe

La mammographie est un examen radiologique des seins qui permet de visualiser avec précision leur structure interne, elle a pour objectif le dépistage du cancer du sein à un stade précoce pour améliorer les chances de guérison. Cette technique permet de recueillir des images grâce à l'utilisation de rayons X. Dans la plupart des cas, 2 clichés sont réalisés sur chaque sein (face et oblique). On distingue généralement deux types de mammographies, la mammographie analogique et la mammographie numérique.

a) La mammographie analogique

La mammographie analogique, aussi appelée "classique" ou "conventionnelle" permet d'obtenir une image imprimée sur un film argentique. Ce type de radiographie est celui qui est le plus couramment utilisé depuis que les campagnes de dépistage du cancer du sein ont été généralisées en 2004. En pratique, lors de l'examen, les patientes posent successivement leurs seins entre deux plaques qui vont les compresser et les aplatir légèrement, permettant ainsi à l'appareil de radiographie de réaliser plusieurs clichés.

b) La mammographie numérique

La mammographie numérique repose sur deux types de technologies. La première permet de transformer les images d'un film argentique en images numériques grâce à un scanner ou une caméra CCD. Ces derniers déterminent une valeur pour toutes les nuances de gris présentes sur les clichés et constituent une image. La seconde catégorie de mammographie numérique, dite "plein champ" ou "numérisation directe", permet quant à elle de construire et restituer une image sur ordinateur grâce à des détecteurs numériques.

En présence de tumeurs, la mammographie permet de distinguer leur caractère malin ou bénin, et grâce à cet examen la mortalité serait diminuée de 20 à 30% chez les femmes de plus de 50 ans, d'après l'organisation mondiale de santé (**OMS**). La mammographie est pratiquée dans les campagnes de dépistage du cancer du sein, et aussi pour le diagnostic, la localisation lors d'interventions chirurgicales [2]. En cas d'anomalie, une biopsie mammaire pourra être

proposée dans certains cas, cet examen médical consiste à prélever un échantillon de la tumeur du sein afin de déterminer si elle est cancéreuse ou non. La mammographie est indiquée à partir de l'âge de 40 ans, tous les deux ans jusqu'à l'âge de 75 ans. Chez les femmes jeunes ou ayant des seins denses, la mammographie est souvent couplée à l'échographie mammaire. La figure I.8 ci-dessous, représente une image mammographique.

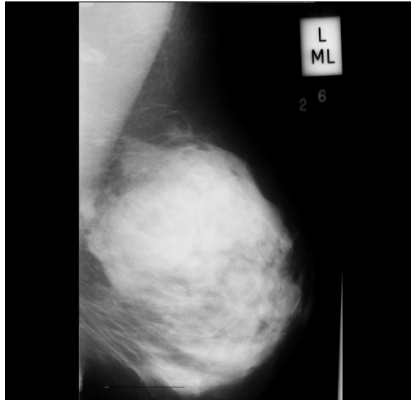


Figure I.8 Résultat d'une mammographie.

I.3.3.2 Limites de la mammographie

La mammographie est le meilleur test dont nous disposons pour détecter le cancer du sein à un stade précoce, mais il n'est pas parfait. Un résultat faux négatif signifie que les résultats de la mammographie sont normaux alors que la femme est en réalité atteinte d'un cancer. La mammographie peut aussi donner un résultat faux positif. Cela signifie que les résultats de la mammographie indiquent que la femme peut être atteinte d'un cancer du sein alors qu'en réalité elle ne l'est pas. D'autres tests montreront que bon nombre de ces anomalies ne constituent pas un cancer.

I.3.3.3 Le mammographe

C'est l'appareil qui réalise les mammographies, il se compose d'un tube radiogène générateur de rayons X de faible énergie et d'un système de compression du sein (Figure I.9). Le sein est placé entre le plateau porte-film et une plaque amovible, afin d'y être comprimé. La radiographie est réalisée sur des films argentiques ou sur des systèmes de radiologie digitale de haute qualité. L'analyse de la glande mammaire est réalisée grâce aux différences de l'atténuation des différents types de tissu.

L'anatomie du sein permet de établir la relation entre la nature du tissu mammaire et l'infiltration des rayons X.

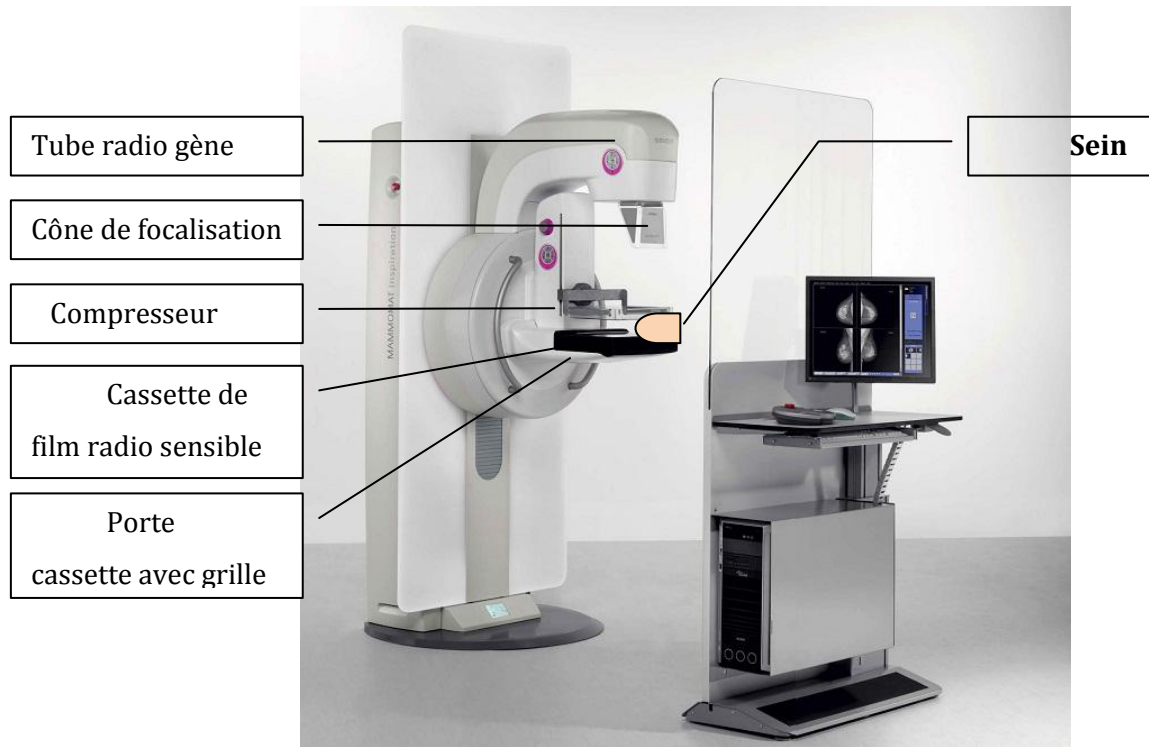


Figure I.9 Les composants d'un mammographe.

I.3.3.3 Relation entre les images de mammographie et l'anatomie du sein

L'atténuation du faisceau de rayons X dépend de type des différents tissus mammaires traversés. En effet, la graisse est considérée comme une zone radio transparente, elle apparaît très sombre sur un cliché mammographique. En revanche, le tissu fibro glandulaire et le calcium (qui est le composant essentiel des lésions mammaires) sont des éléments radio opaques qui apparaissent claires sur un cliché de mammographie.

D'après les informations relatives à l'anatomie et à la radio transparence, on peut affirmer que l'aspect général d'une mammographie est sombre tandis que les zones contenant des microcalcifications ou des masses (composées de calcium) sont plus claires [2].

I.3.3.4 Les incidences en mammographie

À cause de la complexité de l'anatomie du sein, la mammographie est généralement réalisée sous différentes directions appelées incidences. L'objectif d'une bonne incidence est de visualiser le maximum de tissu mammaire en l'étalant le plus possible sur la plaque radiographique.

Les incidences les plus utilisées sont :

1- L'incidence de face ou cranio-caudale ne permet pas d'explorer le sein dans sa totalité. Le porte cassette (le détecteur) ainsi que le tube radiogène sont horizontaux. Cette incidence explore bien la région centrale et la région interne du sein.

2- L'incidence medio latérale oblique (MLO) (encore appelée oblique externe) c'est la plus importante car elle prend en partie le muscle pectoral et permet dans ces conditions d'explorer la quasi-totalité de la glande mammaire sur une seule incidence.

3- L'incidence de profil le porte-cassette est vertical et placé entre les deux seins (profil interne). Cette incidence, pareillement à celle de face, explore la région centrale et interne du sein. La Figure I.10 illustre le positionnement du tube radiogène et du détecteur pour les différentes incidences.

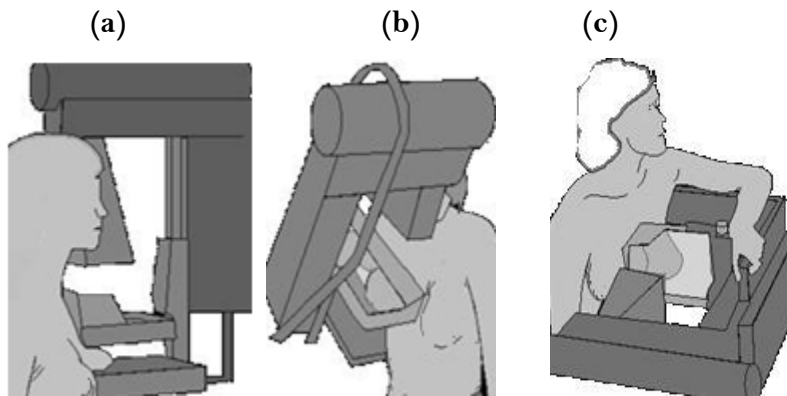


Figure I.10 Illustrations des incidences principales en mammographie

(a) incidence cranio caudale (CC), (b) incidence medio latérale oblique (MLO) et (c) incidence de profil.

1.3.3.5 Les examens mammographiques

La mammographie est actuellement pratiquée dans deux circonstances: dans le cadre d'un dépistage ou d'un diagnostic.

1.3.3.5.1 Le dépistage

Le dépistage du cancer recherche la présence d'une lésion chez des personnes qui n'ont aucun symptôme. Le Centre International de Recherche sur le Cancer (CIRC) confirme l'efficacité du dépistage mammographique systématique pour réduire la mortalité par cancer du sein [4]. Le dépistage a des avantages majeurs puisqu'il permet de diminuer le nombre de décès par cancer. En effet, s'il est détecté tôt, le cancer du sein peut être guéri dans neuf cas sur dix, ce qui participe à une réduction de 15 à 21% du taux de mortalité liée à ce cancer [3].

1.3.3.5.2 Le diagnostic

La mammographie diagnostique est effectuée chez des personnes qui ont des symptômes : masse palpée, région suspecte, etc. Elle est généralement réalisée après un test de dépistage.

L'objectif principal de cette mammographie de diagnostic est soit la recherche d'un signe radiologique dans une zone suspecte, soit l'analyse d'une lésion détectée cliniquement.

Le radiologiste examinera les clichés et recherchera des changements ou des anomalies.

Dans le cas d'une lésion bénigne, un suivi à moyen terme est recommandé. Par contre, dans le cas d'une lésion maligne, le médecin dispose d'autres méthodes d'investigation : les prélèvements de tissu mammaire situés au niveau de la zone suspecte, une chimiothérapie, une radiothérapie ou une hormonothérapie, etc.

1.4 Les différentes lésions mammaires

La généralisation du dépistage du cancer du sein, amène à découvrir beaucoup d'anomalies purement radiologiques. Parmi ces anomalies, on distingue trois grands types de lésions évocatrices d'un cancer:

- Les calcifications mammaires
- Masse (Une opacité irrégulière blanche sur la radiographie).
- La modification de l'architecture normale de la glande mammaire.

I.4.1 Les calcifications

Les calcifications mammaires sont des dépôts de calcium qui se forment dans le tissu du sein. Elles sont associées à des affections bénignes comme par exemple une ancienne plaie, infection ou un kyste calcif [5]. La plupart des calcifications ne sont pas associées au cancer. Afin de s'en assurer, le radiologue étudie leur taille, leur forme et leur disposition à l'aide d'une mammographie sur laquelle elles apparaissent souvent sous forme de petits points blancs. Certaines de leurs caractéristiques, comme une forme irrégulière ou certains regroupements, peuvent être suspectes. Il existe deux types de calcifications : les macrocalcifications et les microcalcifications.

I.4.1.1 Micro calcification

Les microcalcifications sont de minuscules dépôts de calcium dans le sein (< 0,5 mm). Leur présence signifie parfois que l'activité de certaines cellules du sein est accrue. Une cellule plus active absorbe en effet davantage de calcium que celle qui l'est moins. Les microcalcifications peuvent faire suspecter un cancer du sein, notamment lorsqu'elles apparaissent isolées ou regroupées en grappes à la mammographie. Si tel est le cas, une biopsie est indiquée afin de vérifier qu'il ne s'agit pas d'un cancer. La majorité des microcalcifications sont bénignes. Le seul signe précoce de cancer que l'on peut observer sur une mammographie est la présence de microcalcifications. Dans 40% à 70% des cas, selon les études, ces microcalcifications sont associées à un processus malin [5].

L'analyse de la morphologie permet le plus souvent de séparer les microcalcifications bénignes et malignes. Les microcalcifications arrondies ou ovales, uniformes dans leur taille et leur forme, sont souvent bénignes. A l'inverse, celles qui sont irrégulières et hétérogènes sont souvent malignes. La figure (Figure I.11) illustre les différents types de microcalcifications.

a) cutanées ou dermiques, b) vasculaires, c) grossières ou coralliformes, d) en bâtonnets, e) rondes, f) à centres clairs, g) en coquille d'œuf ou pariétales, h) à type de lait clacique, i) de suture, j) dystrophiques, k) amorphes ou indistinctes, l) fines polymorphes, m) linéaires ramifiées [8].

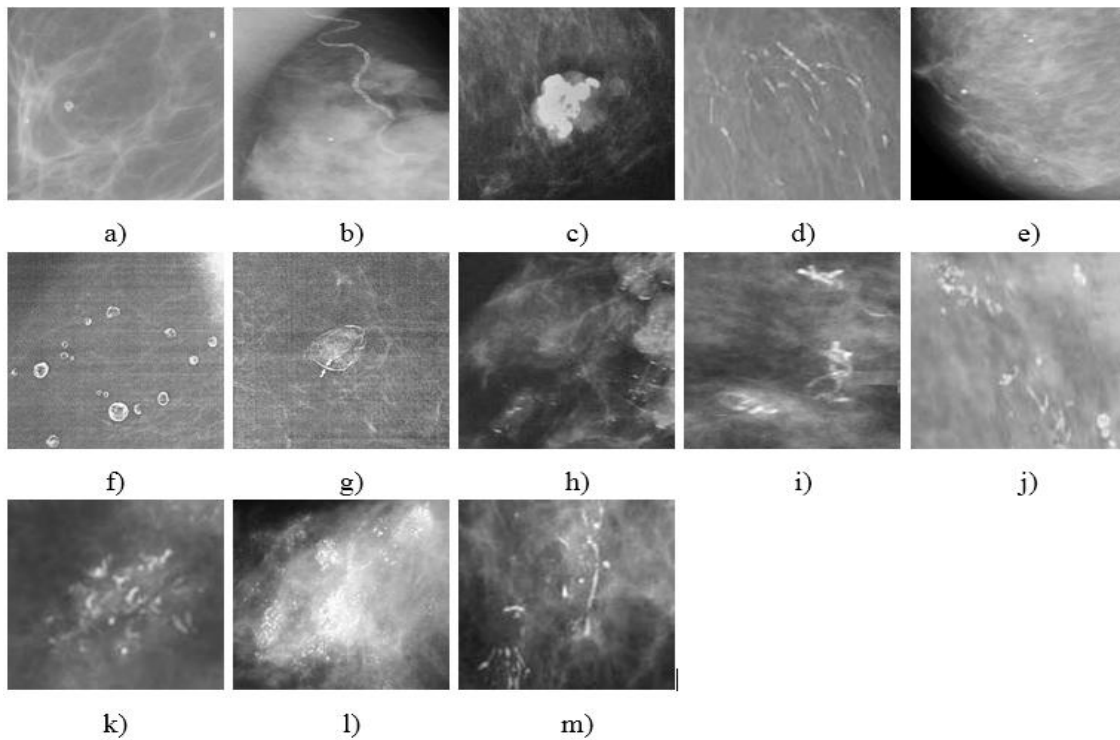


Figure I.11 Les différents types de microcalcifications .

I.4.1.2 Les macrocalcifications

Sont de gros dépôts grossiers de calcium dans le sein. Elles sont plus fréquentes chez les femmes âgées de plus de 50 ans. On les associe souvent à des changements bénins qui se produisent dans le sein et qui sont provoqués par exemple au vieillissement des artères du sein, à d'anciennes lésions, à une infection des tissus du sein, à des masses telles qu'un fibroadénome ou des kystes. Le radiologue ne recommande pas systématiquement de biopsie lorsque ces macrocalcifications sont découvertes.

I.4.2 Les masses

Par définition, une masse est une opacité importante occupant un espace dans le sein et vue comme une tâche blanche sur l'image mammographique. Il peut s'agir d'un kyste ou d'une lésion solide, qui peut correspondre à un cancer de sein. Elle vue sur deux incidences différentes. Dans le cas ou, la masse est observée uniquement sur une seule incidence, elle est appelée asymétrie jusqu'à ce que sa nature tridimensionnelle soit confirmée. Différents attributs permettent aux médecins de la décrire en vue de déterminer leur nature : sa forme, son contour et sa densité.

I.4.2.1 La forme

D'après la classification du BIRADS de l'ACR [6], les masses mammaires peuvent avoir la forme ronde, ovale, lobulée ou irrégulière (Figure I.12).

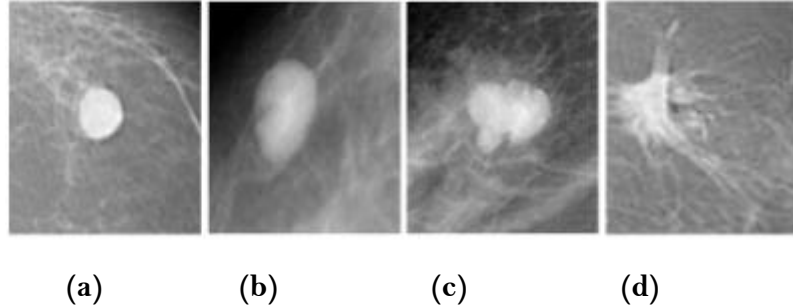


Figure I.12 Les différentes formes possibles d'une masse.

(a) ronde, (b) ovale, (c) lobulaire et (d) irrégulière.

I.4.2.2 Le contour

Le BIRADS dénombre cinq classes de contour que sont les contours : circonscrit, masqué, micro-lobulé, mal-défini et spicule (La Figure I.13).

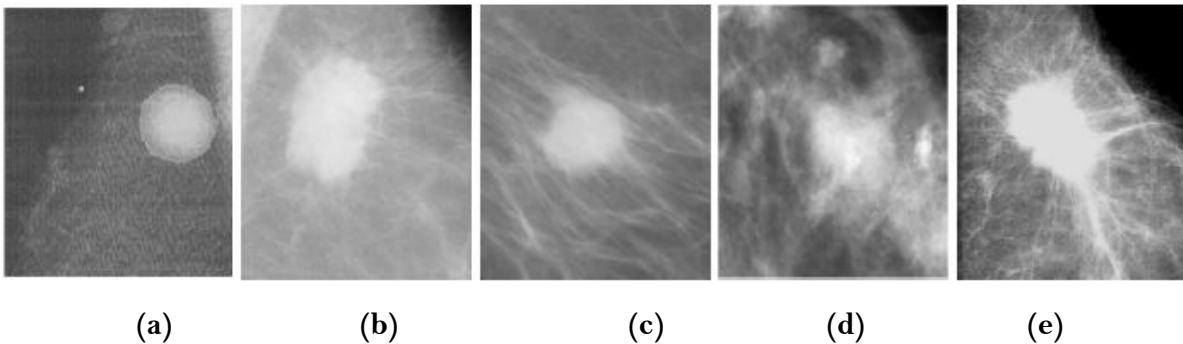


Figure I.13 Les différents types de contour d'une masse:

(a) circonscrit, (b) masqué, (c) micro-lobulé, (d) mal-défini et (e) spicule.

I.4.3 Densité mammaire (DM)

Le sein est composé de différents types de tissus : le tissu glandulaire et conjonctif et la graisse. La densité se définit comme la présence plus ou moins importante d'éléments glandulaires et fibreux par rapport au tissu adipeux. La lecture de la mammographie devient plus difficile (diminution de la sensibilité) à mesure que la DM augmente. Des recherches

récentes ont montré que la densité mammaire est un facteur de risque important, en raison de ce que les pathologies peuvent être cachées à l'intérieur d'une région très dense [9].

Il y a une corrélation entre l'augmentation de la densité mammaire et le risque de cancer du sein. Il est multiplié par 2 à 5 selon la densité [10].

Selon BI-RADS (Breast Imaging Reporting and Data System) de l'American College of Radiology, il existe quatre catégories qui s'expriment en % de tissu glandulaire (Tableau I.1).

Tableau I.1 classification BIRADS.

Catégorie BI-RADS*	DM exprimée en % de tissu glandulaire	
A	< 25%	Seins presque entièrement graisseux
B Majorité des femmes	25 à 49 %	Seins composés de zones de densités fibro-glandulaires éparses
C	50 à 74 %	Seins denses de façon hétérogène, pouvant masquer des petites masses
D Risque augmenté jusqu'à 2 fois par rapport à B	≥ 75 %	Seins extrêmement denses, diminuant la sensibilité de la mammographie

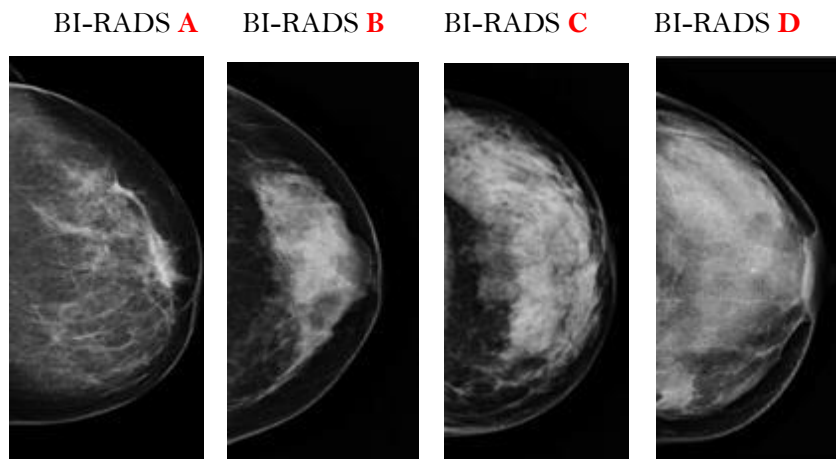


Figure I.14 Différents types de densité mammaire.

I.4.4 La classification des anomalies mammographiques

La classification des résultats d'une mammographie comprend 6 catégories. Ces dernières ont été définies par l'American College of Radiology (ACR) et sont utilisées par la majorité des professionnels de santé à travers le monde.

Les données obtenues à partir d'un examen mammographique peuvent ainsi être classées en :

- **ACR1** : cela signifie que la mammographie n'a révélé aucun problème de santé, la poitrine de la patiente est parfaitement normale.
- **ACR2** : l'examen mammographique a révélé des anomalies bénignes. Sans danger, elles ne nécessitent pas d'examens complémentaires ni de surveillance particulière.
- **ACR3** : cette classification indique la présence d'une anomalie d'apparence bénigne, mais nécessitant tout de même une surveillance à court terme, d'une durée de 3 à 6 mois. La surveillance pourra prendre la forme de nouveaux examens radiologiques, selon une échéance définie par le radiologue et en association avec le médecin traitant.
- **ACR4** : cette catégorie signifie qu'une lésion douteuse ou indéterminée a été détectée et qu'une biopsie et/ou une cytoponction sont nécessaires pour apporter un diagnostic précis. La lésion identifiée peut être bénigne, précancéreuse ou indiquer que la patiente souffre d'un cancer.
- **ACR5** : l'examen mammographique a relevé la présence d'une lésion évocatrice d'un cancer. Les patientes se voient systématiquement prescrire une biopsie ou une exérèse, une intervention chirurgicale servant à retirer un élément nuisible ou inutile à l'organisme, comme une tumeur ou un organe.
- **ACR0** : un tel classement indique que les résultats nécessitent un examen complémentaire, par exemple une lecture par un autre médecin. De nouveaux examens peuvent également être prescrits.

La classification ACR des résultats de mammographies est un outil particulièrement utile permettant d'évaluer l'état de santé d'une patiente et contribue à améliorer la prévention du cancer du sein.

I.5 Diagnostic Assisté par Ordinateur (CAD)

Un système automatique de Diagnostic Assisté par Ordinateur sert à donner un second avis au radiologue, et permet d'aider le radiologue à diminuer le nombre de biopsies inutiles. Un tel système est composé généralement par 4 étapes: un prétraitement qui facilite la tâche de l'étape suivante qui consiste à détecter tous types d'anomalies. Ensuite, une étape d'extraction des descripteurs qui décrivent aux mieux les spécificités de ces lésions est effectuée. Finalement, une étape de classification et de décision. Ces différentes étapes sont plus détaillées dans le chapitre suivant.

I.6 Conclusion

Dans ce chapitre, la notion du cancer du sein ainsi que les outils de l'imagerie médicale permettant le diagnostic de ce type de cancer ont été présentés. Une grande attention a été consacrée à l'étude des différents types des pathologies mammaires.

L'étude des différentes lésions mammaires et leur classification montre, d'une part, la complexité de l'étape de détection des lésions et, d'autre part, la complexité de l'étape de décision (mammographie bénigne ou maligne). Alors, dans ces deux cas, la prise de décision par le radiologue est une tâche suffisamment délicate et primordiale pour la survie ou non d'une femme. Pour cela, plusieurs études de recherche ont été orientées vers l'automatisation de la lecture des mammographies et de la prise de décision.

Chapitre II

Systeme d'aide au diagnostic (CAD)

CHAPITRE II Système d'aide au diagnostic

II.1 Introduction

En réalité, selon les estimations, entre 10 et 30% des cancers du sein ne sont pas détectés par les radiologues lors du dépistage systématique [12]. En outre, la mammographie elle-même ne peut pas prouver qu'une zone suspecte est maligne ou bénigne. Pour décider cela, le tissu doit être retiré pour examen en utilisant des techniques de biopsie du sein. La pénalité des erreurs de détection ou de classification est très élevée. Une détection de faux positif peut entraîner une biopsie inutile. Les statistiques montrent que seulement 20% à 30% des cas de biopsie du sein sont cancéreux. En cas de détection de faux négatifs, une tumeur réelle reste non détectée, ce qui pourrait entraîner des coûts plus élevés, voire le coût d'une vie humaine.

Récemment, le diagnostic assisté par ordinateur (CAD) est devenu une partie intégrante du travail clinique routine pour la détection du cancer du sein sur des mammographies. Cela semble indiquer que la CAD commence à être largement appliqué dans la détection et le diagnostic différentiel de nombreux types différents d'anomalies dans les images médicales. En fait, le CAD est devenu l'un des principaux sujets de recherche en imagerie médicale et en radiologie diagnostique.

Historiquement, c'est à partir de 1950 que l'utilisation des ordinateurs a été considérée pour l'analyse d'anormalités radiographiques. Ce n'est néanmoins qu'à partir de 1980 que les chercheurs et radiologues ont utilisé l'ordinateur comme un système d'aide au diagnostic plutôt qu'un système totalement automatisé [13].

Dans ce chapitre, la motivation et la nécessité du développement des systèmes de CAD sont présentées.

II.2 Définition

Un système d'aide à la décision dédiée à l'imagerie médicale est un système informatique destiné à faciliter la détection et / ou le diagnostic des lésions par le biais d'un "deuxième avis" [14]. Le but des systèmes de (CAD) est d'améliorer la précision de radiologue et de lui donner un second avis objectif sur l'assistance à l'interprétation des images médicales et au diagnostic. Les systèmes de CAD sont classés en deux groupes: les systèmes d'aide à la détection (computer aided detection system CADe) et les systèmes d'aide au diagnostic (computer aided diagnosis system CAD) (cf. figure II.1). Les (CADe) sont des systèmes conçus pour localiser les lésions sur les images médicales. En générale un système de détection automatique de cancers en mammographie se compose de deux branches : une dédiée à la détection des micro-calcifications et l'autre à la détection des masses. Tandis que, les (CAD) effectuent la distinction entre tumeurs bénignes et malignes. Les CAD ce sont le fruit de plusieurs disciplines telles que le traitement d'images, l'intelligence artificielle, la médecine, la biologie,...etc.

Dans notre thèse on s'intéresse aux systèmes d'aide au diagnostic (CAD).

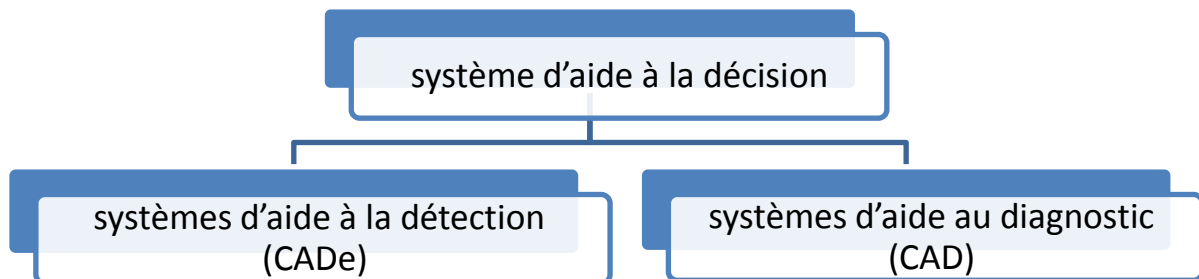


Figure II.1 système d'aide à la décision.

II.3. Etat de l'art des CAD pour le cancer du sein

Depuis quelques années, de nombreux chercheurs ont fait des efforts considérables pour aider les radiologues à détecter et à diagnostiquer les masses cancéreuses en mettant au point des outils d'interprétation assistée par ordinateur pour la mammographie. Le traitement d'images et les systèmes intelligents sont deux courants principaux des technologies informatiques qui ont été explorés en permanence dans le développement de systèmes de mammographie assistée par ordinateur.

Plusieurs chercheurs ont montré que les CAD améliorent les performances de diagnostic des radiologues [15,16].

Récemment, plusieurs techniques ont été développées pour la détection et le diagnostic du cancer du sein. Zakeri et al. [17] ont utilisé les caractéristiques de forme et de texture pour la classification des mammographies en deux classes bénignes et malignes, ils ont atteint une précision de 95,00% avec le classifieur SVM (*support vector machine*). Lahmiri et al. ont utilisé le SVM avec primitives de la DCT afin de classer les mammographies en images normales ou anormale avec une précision d'environ 92,98% [18]. Dans [19], les auteurs ont proposé une méthode de classification par mammographie en tissus bénins, malins et normaux en utilisant une analyse en composantes indépendantes et des réseaux de neurones multicouches. En outre, dans [20], des taux d'environ 86,85% sont atteints par Martins et al. pour classer les régions de mammographie en bénignes, malignes et normales en utilisant un réseau de neurones bayésien. Dans [21], les auteurs ont utilisé des transformées en ondelettes et curvelet pour analyser la micro-classification des images de mammographie et ont obtenu une précision de 93,86% et 90,43%, respectivement.

Un certain nombre de méthodes ont été utilisées pour classifier et / ou détecter les anomalies dans les images médicales, telles que les méthodes statistiques [22] et la plupart d'entre elles utilisaient des caractéristiques extraites à l'aide de techniques de traitement d'images [23]. En outre, d'autres méthodes ont été présentées dans la littérature sur la base de la théorie de logique flous [24], des modèles de Markov [25] et des réseaux de neurones [26]. La plupart des méthodes de CAD ce sont des puissants outils pouvant aider les radiologistes et améliorer les résultats du diagnostic d'un patient. Le système CAD est utilisé pour la détection précoce d'anomalies [27].

Plusieurs travaux ont été réalisés pour développer des outils de détection et de diagnostic du cancer du sein assistés par ordinateur. Dans [28], les auteurs ont proposé une méthode de discrimination et de classification des mammographies de tissus bénins, malins et normaux en utilisant une analyse en composantes indépendantes et des réseaux de neurones multicouches. Une machine à vecteurs de support associée à une analyse en composantes indépendantes a été utilisée dans [29] pour classer les tissus de mammographie.

II.4 Les étapes de développement d'un système d'aide au diagnostique

En pratique, le système d'aide au diagnostique (CAD) en utilisant l'image mammographique passe par plusieurs étapes, qui doivent être exécutées l'une après l'autre. Ses étapes sont :

- Prétraitement qui sert à améliorer la qualité de l'image.
- Segmentation qui permet de supprimer les régions non intéressantes.
- Extraction et sélection des caractéristiques.
- Classification et prise de décision.

Ces différentes étapes sont résumées dans le diagramme représenté dans la figure II.2.

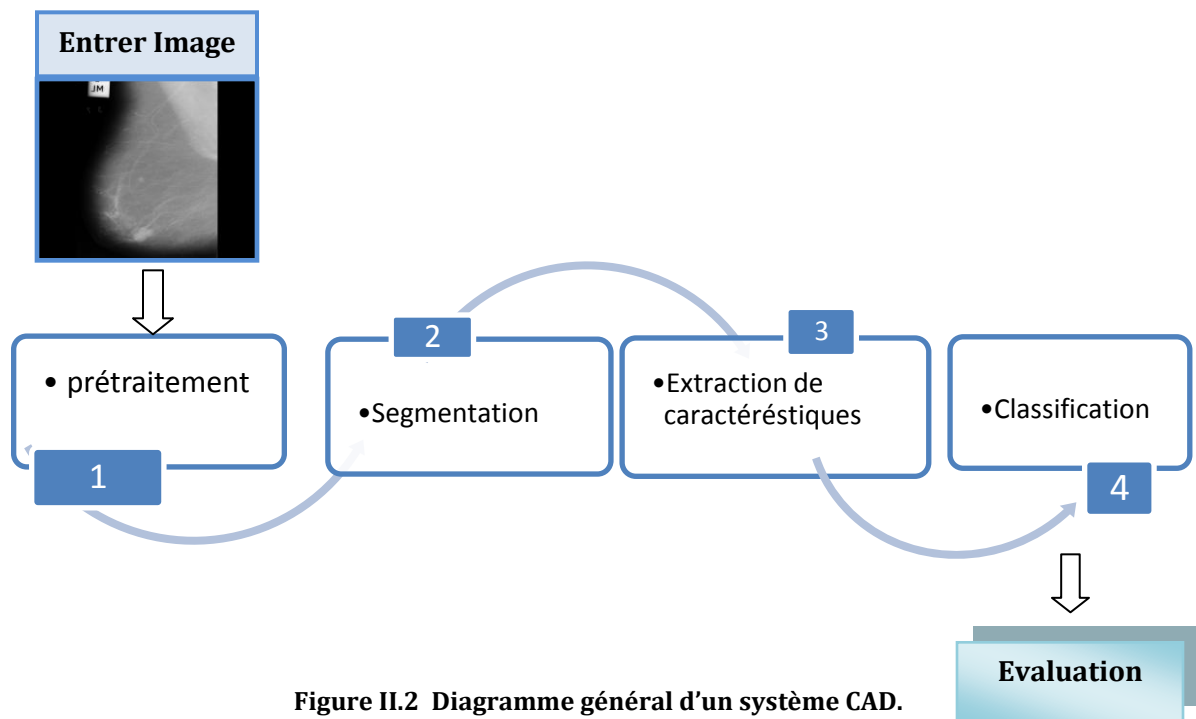


Figure II.2 Diagramme général d'un système CAD.

II.4.1 Prétraitement des images

L'une des difficultés majeures de la conception de CAD robustes est l'absence de normes de qualité d'image. Cependant, le prétraitement souvent utilisée pour améliorer la qualité de l'image. Cela comprend la suppression du bruit, l'élimination des artefacts, les étiquettes ...etc., la sélection de la zone d'intérêt (cf. Figure II.3).

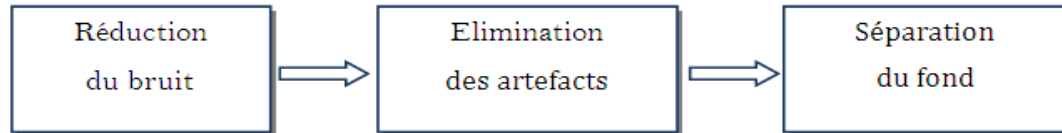


Figure II.3 Système de prétraitement.

II.4.1.1 La réduction du bruit

II.4.1.1.1 Définition

La réduction de bruit (filtrage) est une technique importante d'amélioration de la qualité de l'image numérique. Le filtrage d'image joue un rôle important dans le traitement d'images pour l'analyse d'images. Le filtrage d'image peut être défini comme la récupération de l'image d'origine impure par le bruit. Parmi le grand nombre d'algorithmes de réduction de bruit, le meilleur doit éliminer complètement le bruit de l'image, tout en préservant les caractéristiques importantes.

II.4.1.1.2 les méthodes de filtrages

Dans les images de mammographie, les méthodes de filtrage spatial et les méthodes de filtrage de transformation de domaine sont les deux approches de base du filtrage d'image.

II.4.1.1.2.1 Filtrage spatial

Le filtrage spatial est un moyen traditionnel d'éliminer le bruit des données d'image. Les filtres non linéaires et linéaires sont les deux types de filtres spatiaux.

A) Filtrage linéaires

Les filtres linéaires sont basés sur la contrainte de linéarité qui produit des signaux d'entrée variant dans le temps pour produire des signaux de sortie. Les filtres linéaires améliorent les images de nombreuses manières : netteté des contours des objets, réduction du bruit aléatoire, correction pour un éclairage inégal, déconvolution pour corriger le flou et le mouvement, etc. Mais le problème principal du filtre linéaire est qu'il peut produire un effet de flou.

a) filtre Moyen

Le filtre moyen est un filtre linéaire simple à moyenne. Dans le filtre moyen, la valeur centrale dans la fenêtre est remplacée par un filtre à fenêtre glissante. Chaque valeur de pixel dans l'image est remplacée par la valeur moyenne de ses voisins, y compris elle-même [30]. Le filtre moyen est principalement utilisé lorsqu'une image est corrompue par le bruit du sel et du poivre. Le filtre moyen modifie les valeurs de bruit et de noir affectées par le bruit pour les rapprocher des valeurs de pixel des valeurs environnantes. Si le bruit dans certaines régions doit être éliminé, le filtre moyen est utilisé. C'est à ce moment-là qu'une partie seulement de l'image doit être traitée que le filtre moyen est utilisé.

Sur la Figure II.4, les valeurs 8, 4, 7, 2, 1, 9, 5, 3 et 6 sont les valeurs de pixel. La moyenne de la valeur de pixel est obtenue en prenant la moyenne de ces valeurs. Ici, la moyenne est 5, donc la valeur centrale est remplacée par 5.

8	4	7
2	1	9
5	3	6

*	*	*
*	5	*
*	*	*

Figure II.4 Le filtre Moyen

b) Filtre de Wiener

Le filtre de Wiener est utilisé pour filtrer le bruit qui a corrompu un signal. Il est basé sur une approche statistique. Pour réduire le bruit, le filtre Wiener requiert les informations spectrales du bruit et du signal d'origine. Les filtres sont conçus en fonction de la réponse en fréquence souhaitée. Ce filtre est principalement utilisé pour éliminer le bruit additif. L'objectif principal du filtre Wiener est de réduire l'erreur quadratique moyenne [30]. Les deux parties du filtre de Wiener sont une partie à filtrage inverse et une partie à lissage des bruits.

La déconvolution et la suppression du bruit avec une opération de compression sont effectuées dans la partie filtrage inverse.

B) Filtres non linéaires

Les filtres sont conçus pour régler les problèmes des filtres linéaires, surtout pour ce qui concerne la mauvaise conservation des contours. Ils sont principalement utilisés pour éliminer le bruit non additif. L'inconvénient principal du filtre linéaire est qu'il produit un effet de flou et une filtration de bruit incomplète lors du filtrage d'une image. Pour surmonter cet inconvénient, des filtres non linéaires sont utilisés.

a) Filtre médian

Les filtres médians ont un principe de fenêtre mobile similaire au filtre moyen. C'est une méthode simple basée sur les statistiques d'ordre. Il est utilisé pour lisser les images. La quantité de variation d'intensité entre un pixel et l'autre pixel est réduite en utilisant un filtre médian. Dans cette méthode de filtrage, la valeur de pixel d'origine est remplacée par la valeur médiane d'un pixel voisin. Pour calculer la valeur médiane, commencez par trier toutes les valeurs de pixel dans l'ordre croissant, puis remplacez le pixel calculé par la valeur de pixel du milieu (cf. Figure II.5). Si le nombre de valeurs de pixels est pair, la moyenne des deux valeurs de pixels du milieu est utilisée pour remplacer [30]. En cas de bruit dépendant du signal, les performances du filtrage médian ne sont pas satisfaisantes.

10	5	20
14	80	11
8	3	22

Figure II.5 filtre Médian (la valeur centrale 80 est remplacée par 11).

II.4.1.1.2.2 Transformation de domaine

On prend comme exemple les transformées en ondelettes

Transformée en Ondelettes

La transformation en ondelettes est l'un des puissants outils de traitement du signal et des images. Cette fonction arithmétique divisera les données en différentes composantes de fréquence [31]. Comme il s'agit d'une transformation multi résolution, elle donne des informations temporelles et fréquentielles. C'est une autre forme de représentation d'une image et cela ne modifie pas le contenu de l'information dans l'image. La représentation d'image non redondante est fournie par une transformée en ondelettes discrète (DWT) qui fournit également de meilleures informations spatiales et spectrales. Pour cette raison, le DWT est davantage attiré par le filtrage d'image. Dans le DWT, le signal passe à travers deux filtres complémentaires qui fournissent une composante d'approximation et de détail du signal. Cette étape est appelée décomposition ou analyse. L'inverse de cette étape est appelé reconstruction ou synthèse. L'étape d'analyse et de synthèse est réalisée par l'opération mathématique transformée en ondelette discrète et transformée en ondelette discrète inverse, respectivement. L'image d'origine est transformée en quatre morceaux lorsque l'application DWT est appliquée.

Les quatre pièces sont normalement étiquetées comme A1, H1, V1 et D1 (Figure II.6).

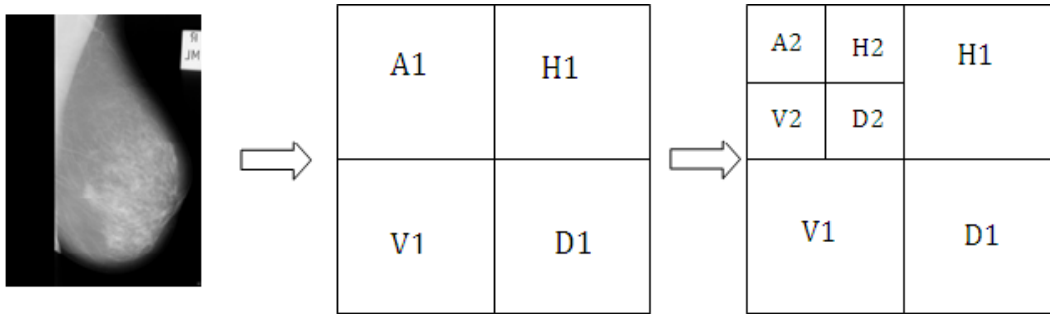


Figure II.6 Décomposition d'ondelettes à différents niveaux (DWT).

L'A1 s'appelle l'approximation, peut être décomposée en quatre sous-bandes. Les bandes restantes sont appelées composants détaillés. Pour obtenir le niveau de décomposition suivant, la sous-bande A1 est encore décomposée [32]. Le principal avantage de la DWT est qu'il s'agit d'une transformation multi résolution. Dans le filtrage d'image, la première image est divisée en différentes sous-bandes, puis des algorithmes de seuillage sont appliqués aux coefficients d'ondelettes. Enfin, on applique une transformation en ondelettes inverse qui donne une image de-bruité. La transformation en ondelettes présente également certains inconvénients. Le manque d'invariance par décalage et le manque de sélectivité directionnelle.

II.4.1.2 Elimination des artefacts

En imagerie médicale, le terme artefact désigne une altération du résultat d'un examen radiologique selon certains procédés techniques utilisés. Ce mot est particulièrement employé pour signaler certains types de dégradation de l'image.

La mammographie contient souvent différents artefacts (cf. Figure II.7, II.8, II.9) sous forme d'étiquettes d'identification, de marqueurs ou encore des régions qui n'appartiennent pas au tissu mammaire tel que le muscle pectoral. Ces artefacts sont classés en deux catégories principales: étiquettes de film et artefacts opaques. [33]. cette étape consiste à éliminer tous artefacts, les étiquettes qui peuvent être présents dans l'image .

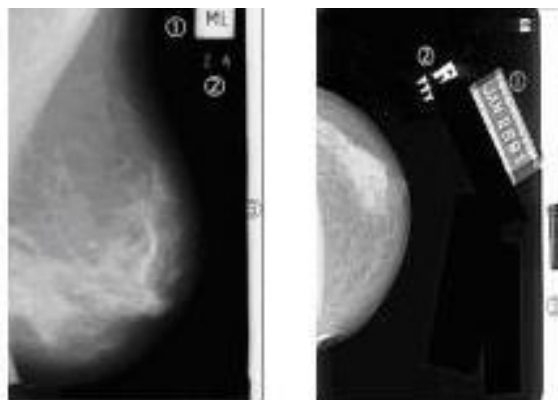


Figure II.7 Exemple d'une image mammographies contient radiopaque artefacts:

1) étiquettes de film. 2) marqueur radiopaque.

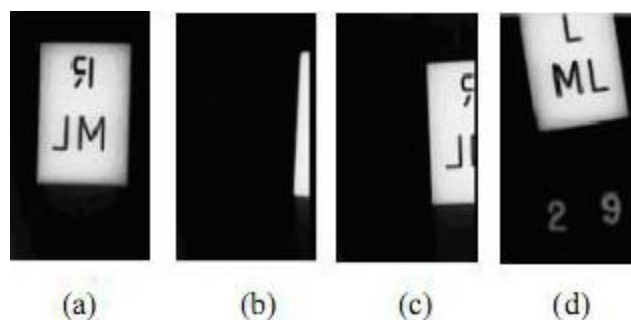


Figure II.8 Exemples d'étiquettes du film mammographique:

(a) label entier et (b-c-d) labels partiels.

Plusieurs travaux sur l'extraction de la région d'intérêt et la suppression d'artefacts en mammographie. Se basent sur la reconstruction morphologique pour éliminer les artefacts d'intensité supérieure.

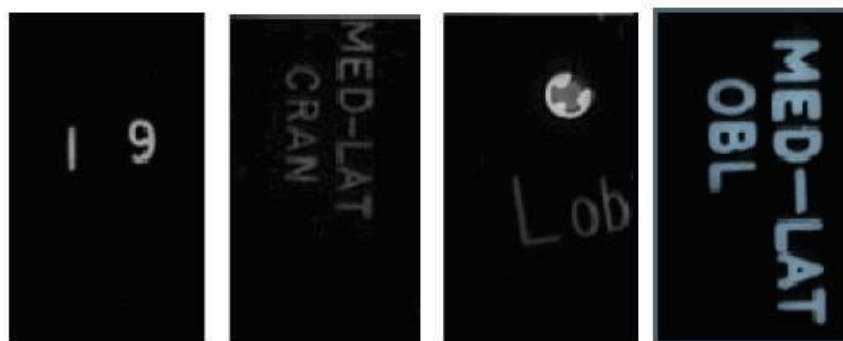


Figure II.9 Exemples de marqueurs opaques sur une mammographie.

II.4.2 Segmentation

La segmentation est une des étapes critiques de l'analyse d'images qui conditionne la qualité des traitements d'images effectués ultérieurement. Elle permet d'isoler dans l'image les objets sur lesquels doit porter l'analyse. En effet, à partir des résultats de la segmentation, il est possible d'identifier les objets d'une image. Par conséquent, les méthodes d'évaluation des algorithmes de segmentation constituent une autre dimension des systèmes de CAD sur le plan pratique.

II.4.2.1 Définition

La segmentation est un traitement de bas-niveau qui consiste à partitionner une image I en sous-ensembles R_i connexes et homogènes (intensité, texture,...), pour pouvoir en extraire les objets qui la composent (régions, contours, ...etc.) [34]. la segmentation d'une image numérique I utilisant un prédicat d'homogénéité P est communément définie comme une partition $I = R_1, R_2, R_3, \dots, R_n$ (R : régions) telle que :

1. $I = \cup R_i, i \in [1 \dots n]$
2. R_i est connexe, $\forall i \in [1 \dots n]$
3. $P(R_i) = \text{vrai}, \forall i \in [1 \dots n]$
4. $P(R_i \cup R_j) = \text{faux} \forall i \neq j$
5. $R_i \cap R_j = \Phi$ (un ensemble vide).

Où le prédicat P est utilisé pour tester l'homogénéité.

II.4.2.2 Les approches de segmentation

Il existe de nombreuses techniques de segmentation, ces méthodes sont couramment classées en trois catégories (Figure II.10) :

- Les approches basées pixel.
- Les approches basées région.
- Les approches basées contour.

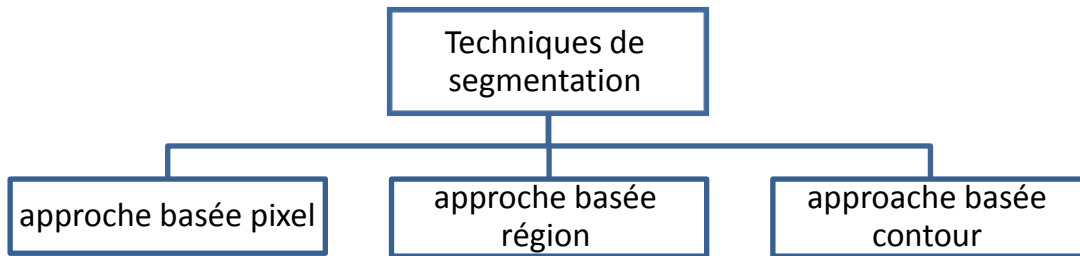


Figure II.10 les techniques de segmentation d'image.

II.4.2.2.1 approche par seuil

Un seuil est une valeur numérique correspondant à un paramètre de l'image (exemple : le niveau de gris), et l'histogramme des niveaux de gris est utilisé afin d'en déduire le seuil [35]. Les régions sont déduites du seuillage. Le seuillage peut être :

- **Global** : un seuil pour toute l'image (cf. Figure II.11)
- **Local** : un seuil pour une portion de l'image.
- **Adaptatif** : Nous avons besoin de séparer l'image en sous images, et de traiter chacune avec son propre seuil.

Le seuillage le plus simple est le seuillage global, il a pour principe :

Soit i l'intensité d'un pixel et f une fonction définissant le niveau de gris et S un seuil:

$$f(i) = \begin{cases} 0, & f(i) < S \\ 1, & f(i) \geq S \end{cases}$$

Le résultat sera évidemment une image binaire.

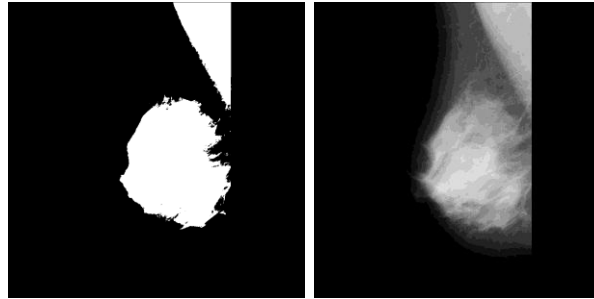


Figure II.11 Exemple seuillage global .

II.4.2.2.2 Approche par régions

La segmentation d'image par l'approche région consiste à regrouper des points selon des propriétés communes. Globalement, elle peut être définie comme une partition d'une image I en une ou plusieurs régions R_1, \dots, R_n telles que : [35]

$$I = \bigcup_{i \in [1..n]} R_i, \quad R_i \cap R_j = \emptyset \quad (\text{ensemble vide}) \quad \forall i \neq j.$$

On distingue deux techniques de cette approche: Croissance de région, Segmentation par division et fusion (Split & Merge).

II.4.2.2.2.1 Méthodes par croissance de régions (SRG)

L'approche de croissance de région est une méthode de segmentation d'image proposée initialement par R. Adams et al. [76] Le processus comprend trois étapes principales: la sélection d'un point de départ (*seed point*), la croissance et la fusion des régions.

Le principe de ces méthodes est de réunir de façon itérative un ensemble de points connectés en une région de plus en plus large, en fonction de critères d'homogénéité. Pour définir une région, on définit un germe dans la région d'intérêt qui sert comme un point de départ pour l'agrégation.

Avantages et inconvenients

Tableau II.1 Avantages et inconvenients du methodes par croissance de regions.

Avantages	Inconvénients
<ul style="list-style-type: none"> ▪ Rapide. ▪ Facile à mettre en œuvre. 	<ul style="list-style-type: none"> ▪ Algorithme très sensible au bruit; ▪ Obtention de frontières non-régulières; ▪ Peu efficace dans le cas de dégradés; ▪ L'ordre du traitement des pixels germes peut avoir une influence sur le résultat; ▪ Choix critique de la valeur des seuils: <ul style="list-style-type: none"> • Risque de sur-segmentation, sous-segmentation; • Dépend du type d'images à traiter

II.4.2.2.2 Segmentation par division-fusion (Split and Merge)

L'algorithme "Split & Merge" a été présenté la première fois en 1974 par Pavlidis et Horowitz [36]. Cet algorithme est composé de deux étapes. Dans la première étape, nous allons utiliser l'opération de (division) en utilisant la structure "quad-tree" pour partitionner l'image en plusieurs régions (**Figure II.12**). Par la suite, une étape de fusion intervient, en regroupant les régions similaires suivant un critère de fusion (**Figure II.13**) Ce procédé est répété jusqu'à l'obtention d'une segmentation [37].

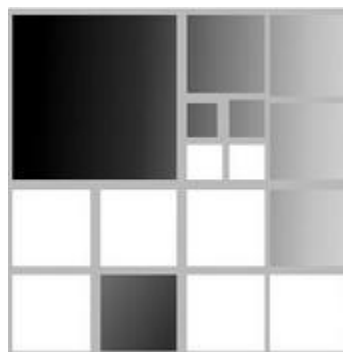


Figure II.12 partitionnement d'image par division en utilisant "quad-tree".



Figure II.13 regroupement des régions similaires suivant un critère de fusion.

❖ Etape de division (split)

- Diviser récursivement tout bloc non homogène selon un prédicat défini: Variance, max-min, ...
- La division d'un bloc donne 4 sous-blocs;
- Les attributs de chaque sous-bloc sont recalculés.

❖ Etape de fusion (merge)

- Regrouper les blocs adjacents représentant des régions homogènes selon un prédicat défini.

II.4.2.2.3 Approche par contour

L'approche basée contour est l'une des premières méthodes de segmentation et jusqu'à présent cette méthode est d'importante utilisation. Un contour est un ensemble des pixels d'une image numérique qui correspond à un Changement brutal de l'intensité lumineuse. [38] Dans l'approche " **contour** ", on considère que les primitives à extraire sont les lignes de contrastes séparant des régions de niveaux de gris différents et relativement homogènes, ou bien des régions de texture différentes. En pratique, il s'agit de reconnaître les zones de transition et de localiser au mieux la frontière entre les régions.

II.4.2.2.4 Comparaison des techniques de segmentation

La Tableau II.2 présente une brève comparaison entre les principales approches de segmentation, concise leurs avantages et leurs inconvénients [39].

Tableau II.2 Comparaison des techniques de segmentation.

	Description	Avantages	Inconvénients
approches basés Région	Regroupement de pixels dans des régions homogènes en fonction de points d'origine	-Concept simple. Il fonctionne bien avec le bruit respect. -Nous pouvons choisir plusieurs critères en même temps	Dépendance aux points de départ sélectionnés. coûteuses pour la mémoire et le temps de calcul.
approches basés seuillage	Atteindre un seuil basé sur les pics et les creux de l'histogramme des images correspondant aux région.	-Fonctionne bien avec une faible complexité de calcul dans la plupart des images. -Calcul Simple pour le seuil.	Ne fonctionne pas bien dans les images à spectre de couleurs proche.
Approches basés contour	Il fonctionne sur l'identification des discontinuités nettes dans l'image.	Facile à percevoir. Fonctionne bien pour une image de bon contraste.	-Très sensible au bruit -Ne fonctionne pas bien sur les images à faible contraste et à changement régulier

II.4.2.3 Segmentation du muscle pectoral

La suppression du muscle pectoral une tâche difficile et très nécessaire, lors du prétraitement de la plupart des vues MLO de mammographies. Cette étape est importante pour obtenir une image sans aucune information supplémentaire en dehors du profil du sein, ce qui donnera ensuite une classification plus précise dans les systèmes de diagnostic (CAD). L'élimination du muscle pectoral est très compliquée à cause des caractéristiques du muscle suivantes [82] :

- Une zone homogène contenant les pixels les plus intenses de la mammographie située dans le coin supérieur gauche /ou droit.
- Le bord du muscle pectoral n'est pas une ligne droite, mais peut être convexe ou concave avec des caractéristiques différentes qui varient d'une image à l'autre.

- Les informations texturales sont très similaires à celles du tissu mammaire.

Malgré ces difficultés, de nombreuses méthodes ont été proposées dans la littérature pour la suppression de muscle pectoral. Les auteurs dans [65, 66, 67] ont proposé une méthode d'ajustement en ligne droite avec la transformation de Hough pour séparer le muscle pectoral et le mammaire, tandis que le travail dans [68] présente un nouvel algorithme utilisant le seuillage optimal et la transformation de Hough pour supprimer le muscle pectoral. Ces approches sont basées sur l'idée que le muscle pectoral a une forme triangulaire dans laquelle elles visent à trouver l'hypoténuse de ce triangle, mais généralement, le bord du muscle pectoral est une courbe plutôt qu'une ligne droite [69]. Jawad et al. Dans [70] ont proposé une autre méthode, ils ont utilisé l'algorithme de croissance de région avec un prétraitement morphologique. Dans [71], Oliver Malagelada a utilisé l'algorithme de croissance de région (SRG) pour localiser le muscle et l'extraire du sein. Dans [72-73], les auteurs ont utilisé l'algorithme k-moyennes adaptatif et l'algorithme de croissance de région. Cependant, il n'existe pas de technique de segmentation standard universelle capable de produire des résultats satisfaisants dans tous les cas. Cependant, il n'existe pas de technique de segmentation standard universelle qui puisse produire des résultats satisfaisants dans tous les cas.

II.4.3 Extraction de caractéristiques

L'extraction de caractéristiques généralement signifie Le calcul des descripteurs de caractéristiques à partir d'une image afin de réduire la dimension de données. Ce sont les caractéristiques de l'ensemble de l'image ou de la région concernée. Le bon choix des caractéristiques a une importante influence sur (1) la taille de la mémoire, (2) la précision de la classification, (3) le coût de la classification et (4) la robustesse.

Selon [40], les descripteurs d'images sont classés en trois catégories: les descripteurs basés sur la forme, les textures et les couleurs (cf. Figure II.14).

On peut en trouver sur la texture comme par exemple, la différence de niveaux de gris [41]. Pour les masses, la première classe de caractéristiques basées sur la forme de l'objet c'est le périmètre par exemple. Des mesures sur l'intensité peuvent aussi être utilisées [42]. On peut par exemple analyser la distribution des niveaux de gris dans la lésion supposée, sur son contour ou dans son voisinage.

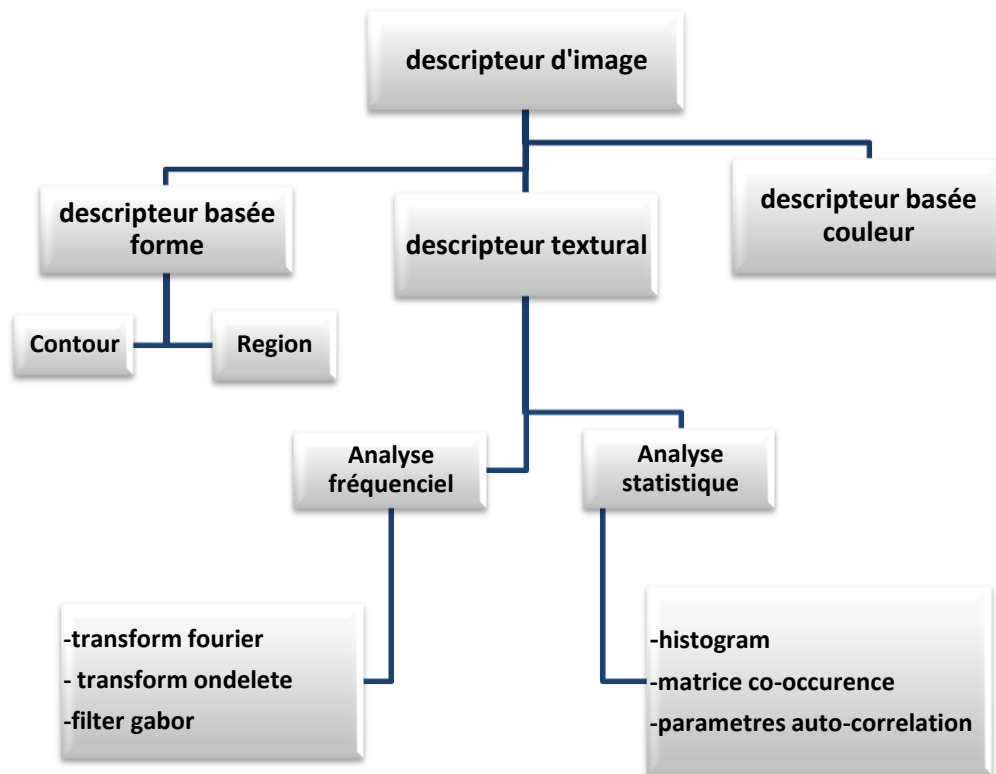


Figure II.14 Vue générale de descripteur d'image.

Récemment, plusieurs chercheurs ont utilisé des caractéristiques texturales pour le développement du système **CAD** du sein. La cooccurrence fonction de texture est l'un des importants caractéristiques qui sont largement utilisés dans la conception de Système de CAD.

Filipczuk et al. [43] ont utilisé la matrice de cooccurrence de niveau de gris (**GLCM**) et la matrice de longueur de passe en niveau de gris (**GLRLM**) pour développer un système automatique de diagnostic du cancer du sein. Islam M. et al. [44] ont introduit une méthode efficace pour la classification maligne-bénigne des mammographies numériques. Les auteurs ont utilisé sept caractéristiques de texture (extraites pour chaque retour sur investissement) comme entrée du classifieur **MLP** (Multi-layer Perceptron) avec 7 unités dans la couche d'entrée, 5 unités dans la couche cachée et 1 unité dans la couche de sortie. Les descripteurs basés sur les couleurs sont des repères visuels pour la récupération d'images et d'objets reconnaissance [45]. Ils sont rarement utilisés dans les travaux du domaine de la détection du cancer du sein.

Les descripteurs de forme sont largement utilisés pour concevoir l'extraction des caractéristiques dans les systèmes de CAD. Au cours des dernières années, l'extraction de caractéristiques de forme a fait d'une attention accrue dans le cas du dépistage du cancer du sein.

Afin de classer les mammographies en malignes ou bénignes, [46] ont proposé un ensemble de caractéristiques de forme pour mesurer la rugosité des contours. La compacité, les moments et le descripteur de Fourier sont extraits de la région d'intérêt et les résultats montrent qu'une plus grande précision est obtenue en utilisant une composition de ces trois facteurs de forme.

II.4.4 Sélection de caractéristiques

Les performances de CAD dépendent plus de l'optimisation de la sélection des caractéristiques que des méthodes de classification. Le grand nombre des caractéristiques extraites souvent, diminue la performance du classificateur.

Des heuristiques comme l'aire sous la courbe **ROC** du classifieur [47], permettent de modéliser la qualité des caractéristiques extraites. Cela permet d'utiliser des méthodes d'optimisation comme les algorithmes génétiques pour obtenir une sélection des caractéristiques optimales. Ce genre d'approche a été mis en œuvre avec succès en mammographie [48].

II.4.5 Classification des anomalies

La classification c'est la dernière étape qui donne la réponse finale sur ce qui est détecté et considéré comme un signe de lésion. L'idée est de combiner les caractéristiques extraites précédemment pour obtenir une décision.

L'étape de classification fait appel aux outils de classification de l'intelligence artificielle, telles que les réseaux de neurones (ANN), les machines à vecteurs de support (SVM), Naïve Bayes (NB), et les méthodes de type plus proches voisins (KNN) ...etc.

Dans le chapitre suivant en détaillant sur les types de classification et les différents outils utilisés.

II.5 Conclusion

Dans ce chapitre nous avons donné un aperçue générale sur les systèmes d'aide aux décisions et leurs différentes parties, ainsi que les techniques utilisées dans le domaine d'imagerie médicale et plus précisément dans les images mammographie.

nous avons cité un état de l'art sur les efforts considérables de nombreux chercheurs pour aider les radiologues en mettant au point des outils d'interprétation assistée par ordinateur pour la mammographie.

nous avons intéressé aux systèmes d'aide au diagnostic, en résumant leurs étapes: le prétraitement et leur rôle dans l'amélioration de la qualité de l'image, la segmentation qui supprime les régions non intéressantes, l'extraction et la sélection des caractéristiques pour réduire la dimension de données, et enfin la classification et la prise de décision.

nous avons résumé les approches de base du filtrage pour les images de mammographie, ainsi les approches de segmentation, leurs avantages et leurs inconvénients. et enfin nous avons cité un état de l'art sur la suppression du muscle pectoral et leur importance dans la précision de classification.

CHAPITRE III

La Classification Des images médicales

CHAPITRE III La classification des images médicales

III.1 Introduction

La classification joue un rôle important dans de nombreux domaines scientifiques et techniques, elle peut être trouvée dans plusieurs domaines d'application comme la reconnaissance des formes, l'analyse discriminante, reconnaissance de la parole, aide au diagnostic, la théorie de la décision.

La classification est la dernière étape des systèmes de CAD permettant de différencier et d'étiqueter l'anomalie.

Les méthodes de classification jouent un rôle important dans le diagnostic et la détection des anomalies. Elles sont classées en deux groupes, comme illustré à la Figure III.1.

Généralement, dans la mise en œuvre du classifieur dans le traitement des images cliniques, des techniques de classification supervisée sont utilisées.

On s'intéresse à la classification des images médicales en général et les images mammographies en particulier telles que les images utilisées dans le domaine de détection et/ou de diagnostic du cancer du sein. On peut distinguer selon le type d'apprentissage entre deux types de classification : la classification supervisée et la classification non supervisée.

Dans ce chapitre, nous allons présenter un panorama des méthodes de classification les plus utilisées dans l'imagerie médicale ainsi que ses principaux aspects.

III.2 Définition

La classification est l'une des techniques les plus anciennes d'analyse et de traitement de données, elle consiste à assigner un objet à un groupe parmi un ensemble de groupes prédéterminés selon un ou plusieurs critère(s) de classification à l'aide d'un classifieur.

Plusieurs définitions ont été proposées:

Michie et al [49] définissent la classification par : "La classification est l'action de regrouper en différentes catégories des objets ayant certains points communs ou faisant partie d'un même concept, sans avoir connaissance de la forme ni de la nature des classes au préalable, on parle alors de problème d'apprentissage non supervisé ou de classification automatique, ou l'action d'affecter des objets à des classes prédéfinies, on parle dans ce cas d'apprentissage supervisé ou de problème d'affectation" .

Pour Mari et Napoli [50]: "Effectuer une classification, c'est mettre en évidence des relations entre des objets, et entre ces derniers et leurs paramètres". Selon Henriet [51] Un problème de classification: "consiste à affecter des objets, des candidats, des actions potentielles à des catégories ou des classes prédéfinies". Nous détaillons, plus particulièrement, les méthodes de la classification supervisée qui font l'objet de nos travaux.

III.3 Les Méthodes de Classification

Généralement, les méthodes de classification ont pour but de regrouper les éléments d'un ensemble $X = \{X_1, \dots, X_n, \dots, X_N\}$ de nature quelconque, en un nombre C optimal de classes selon leurs ressemblances [52]

Parmi les méthodes de classification, on distingue (cf. Figure III.1):

- La classification supervisée (avec apprentissage) dans laquelle chaque image est associée à une étiquette qui décrit sa classe d'appartenance, dans ce cas, un expert identifie les classes de données et on classe les données existantes parmi elles.
- La classification non supervisée (sans apprentissage) dans laquelle le nombre de classes, inconnu a priori, est déduit directement des données. les données disponibles ne possèdent pas d'étiquettes ; il appartient alors au système d'extraire une règle d'appartenance de chaque image à un groupe donné. Dans ce cas, les classes sont construites en fonction des données, selon l'algorithme de classification employé.

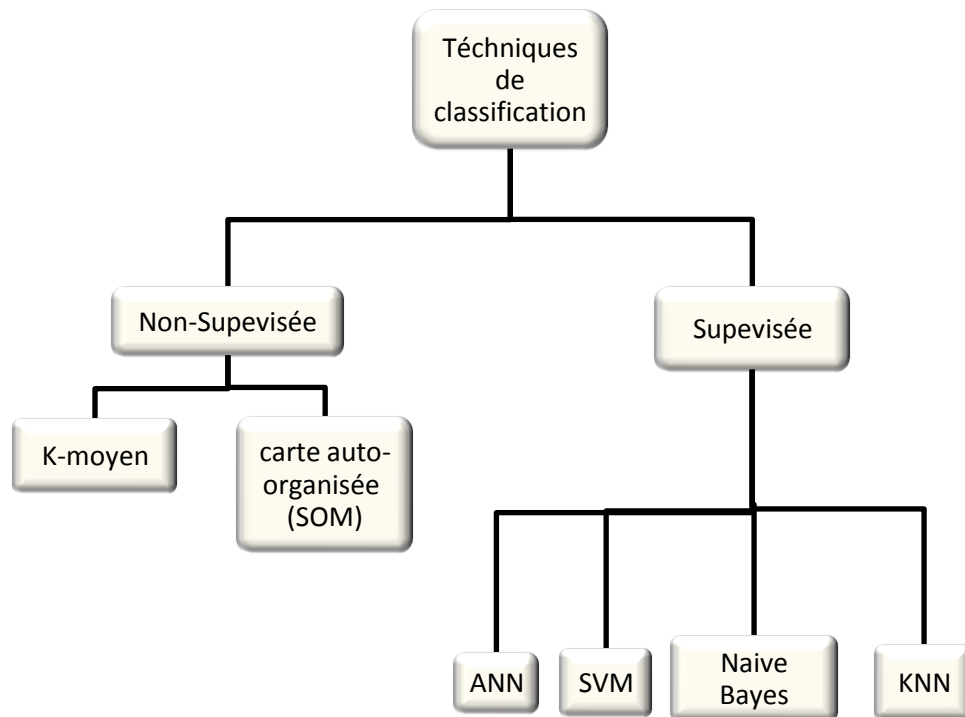


Figure III.1 Différents techniques de classification.

III.3.1 Méthodes de Classification supervisées

Ces sont des méthodes dans lesquelles les classes sont connues *a priori*, chaque image est associée à une étiquette qui décrit sa classe d'appartenance, dans ce cas, un expert identifie les classes de données et on classe les données existantes parmi elles. Elles demandent une phase d'apprentissage dans le but d'apprendre les caractéristiques de chaque classe et une autre phase de test pour décider l'appartenance d'un individu à telle ou telle classe.

Nous pouvons donc citer des algorithmes de classification supervisés de la littérature. Nous présentons l'algorithme "Support Vector Machine" ou "Séparateur à Vaste Marge" (SVM), Les K-plus proches voisins (KNN), Naïve Bayes et enfin les réseaux de Neurones artificiels (ANN).

III.3.1.1 K-Plus Proches Voisins (KNN)

Le classifieur k-plus proches voisins est un algorithme statistique simple en apprentissage automatique, mais un outil puissant. Il a été utilisé dans l'estimation statistique, la reconnaissance de formes et en particulier dans les tâches de classification. Dans la version la plus simple, k-NN prend un nombre arbitraire (k) de voisins plus proche de l'ensemble d'apprentissage et, pour chaque point de test, nous commençons par déterminer tous ses k -voisins les plus proches parmi les points d'apprentissage. La classe que nous assignons au nouveau point est alors la plus fréquente. Le principe de cet méthode consiste à calculer la distance d'un nouvel objet par rapport à ceux dont la classe est déjà connue; le nouvel objet appartiendra à la classe dont il est le plus proche [53]. Dans un contexte de classification d'une nouvelle observation \mathbf{x} , l'idée fondatrice est de faire voter les plus proches voisins de cette observation. La classe de \mathbf{x} est déterminée en fonction de la classe majoritaire parmi les k plus proches voisins de l'observation \mathbf{x} . La méthode k -NN est donc une méthode basée sur le voisinage, non-paramétrique [54].

La méthode k -NN est une approche basée sur des mesures de distance. Il n'a pas besoin d'apprentissage. Lors de la phase de décision, l'algorithme évalue la classe majoritaire des K objets les plus proches de l'exemple à tester [55].

en classification k -NN, le résultat est une classe d'appartenance. Un objet d'entrée est classifié selon le résultat majoritaire des statistiques de classes d'appartenance de ses k plus proches voisins, (k est un nombre entier positif généralement petit). Si $k = 1$, alors l'objet est assigné à la classe d'appartenance de son proche voisin. La Figure III.2 montre un exemple simple d'une classification par l'algorithme KNN avec un calcul de distance euclidienne : le nouvel exemple est associé à la classe B car la majorité de ses plus proches voisins appartiennent à la classe B.

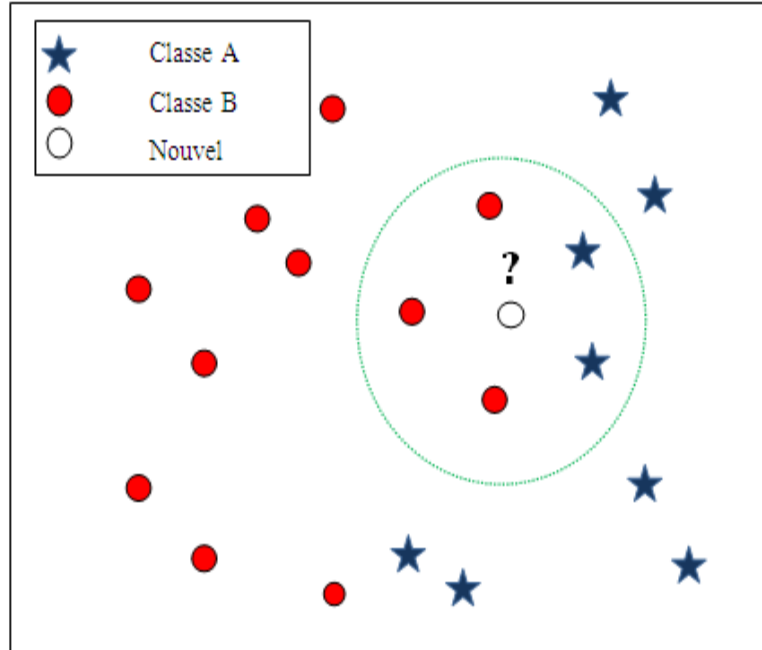


Figure III.2 illustration simple par un algorithme KNN (K=5).

Les 5 points les plus proches du point (?) que l'on cherche à classer.

III.3.1.1.1 Calcul de similarité dans l'algorithme K-NN

L'algorithme K-NN a besoin d'une fonction de calcul de distance entre deux observations. Plus deux points sont proches l'un de l'autre, plus ils sont similaires et vice versa.

Il existe plusieurs fonctions de calcul de distance, notamment, la distance euclidienne, la distance de Manhattan, la distance de Minkowski, celle de Jaccard, la distance de Hamming...etc. On choisit la fonction de distance en fonction des types de données qu'on manipule. Voici quelque définition mathématique des distances.

a) La distance euclidienne

- distance qui calcule la racine carrée de la somme des différences carrées entre les coordonnées de deux points :

$$d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad \text{distance euclidienne (3.1)}$$

b) Distance Manhattan

- la distance de Manhattan: calcule la somme des valeurs absolues des différences entre les coordonnées de deux points :

$$d(x, y) = \sum_{i=1}^k |x_i - y_i| \quad \text{distance Manhattan} \quad (3.2)$$

c) Distance Hamming

- la distance entre deux points donnés est la différence maximale entre leurs coordonnées sur une dimension.

$$d(x, y) = \sum_{i=1}^k |x_i - y_i| \quad (3.3)$$

$$\text{avec} \quad \begin{cases} x=y \rightarrow d=0 \\ x \neq y \rightarrow d=1. \end{cases}$$

d) Distance Chebyshev

$$d(x, y) = \text{Max}(|x_i - y_i|) \quad \text{distance Chebyshev} \quad (3.4).$$

x_i échantillon de test avec k caractéristiques, y_i échantillons d'apprentissage avec k caractéristiques. Notez bien qu'il existe d'autres distances selon le cas d'utilisation de l'algorithme, mais la distance euclidienne reste la plus utilisée.

III.3.1.1.2 Comment choisir la valeur K

K représente un paramètre important de l'approche car il conditionne la façon dont le classifieur va s'adapter aux données en présence, la résolution du classifieur et sa généralité.

Le choix de la valeur K , varie en fonction du de données. généralement, moins on utilisera de voisins plus on sera sujette au sous apprentissage et on obtient un classifieur dit de bonne résolution. Par ailleurs, plus on utilise un nombre K grand, on obtient un classifieur peu sensible au bruit, avec une variance faible mais un biais élevé [56].

III.3.1.1.3 Limitations de K-NN

K-NN est un algorithme assez simple d'un point de vue conceptuel. Principalement, grâce au fait qu'il n'a pas besoin de modèle pour pouvoir effectuer une prédiction. Malgré sa simplicité de mise en œuvre, les performances de l'algorithme K-NN sont encore mal connues et les coûts sont importants en stockage et temps de classification. En pratique, cet algorithme est assez peu utilisé dans sa forme première car coûteux en puissance de calculs. En outre, le choix du type de la distance ainsi que le nombre de voisins K peut ne pas être évident. Il faut essayer plusieurs combinaisons et faire du tuning de l'algorithme pour avoir un résultat satisfaisant.

III.3.1.2 Séparateur à Vaste Marge (SVM)

Le "Support Vector Machine" ou "Séparateur à Vaste Marge" (SVM) [57] est un classifieur discriminant paramétrique établissant un hyperplan séparateur de marge maximale entre les exemples représentants de chacune des classes des données d'apprentissage.

Il est maintenant largement appliqué dans de nombreux domaines tels que la classification, la reconnaissance faciale et le diagnostic du cancer du sein [77]. SVM est un classifieur linéaire binaire qui, pour chaque donnée d'entrée donnée, prédit une classe spécifique à un nouveau vecteur de caractéristiques inconnu. L'idée de base de SVM est de trouver un hyperplan optimal qui sépare parfaitement une donnée en deux classes, en maximisant la taille de la marge. Pour les caractéristiques vectorielles d'un ensemble d'apprentissage, un classifieur SVM construit un hyperplan, qui les sépare en deux classes distinctes: maligne ou bénigne. Pour les données séparables linéairement, un hyperplan est donné comme suit [78]:

$$g(x) = w^t x + b = \sum_{i=1}^n w_i^t x_i + b \quad (3.5)$$

Où x est le vecteur d'entrée, b est un scalaire et est (n) un vecteur normal dimensionnel de cet hyperplan. Cette technique, également appelée méthode des séparateurs à vastes marges, repose sur deux idées clés : la notion de marge maximale et la notion de noyau [60].

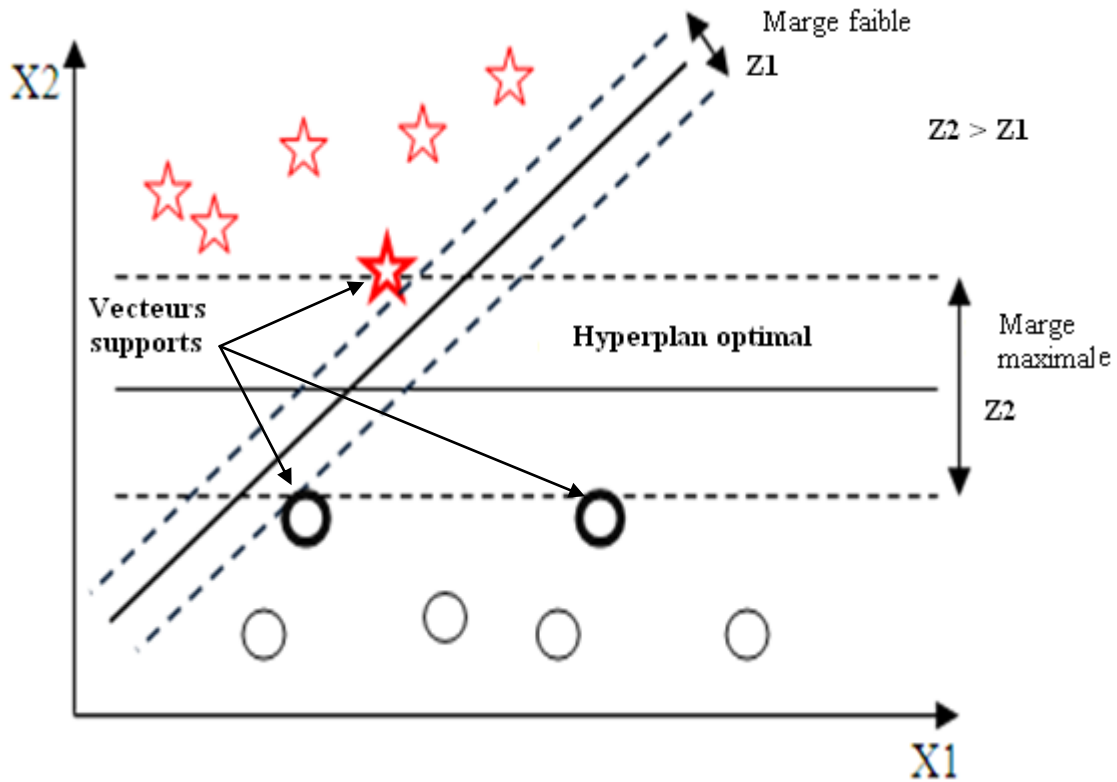


Figure III.3 Représentation de l'hyperplan optimal pour des données séparables.

La figure (figure III.3), illustre un problème de classification binaire en 2 dimensions. IL exist une infinité de droites séparatrices des deux classes en présence mais une seule maximise la distance aux exemples aux frontières des classes (i.e. maximise la distance Z_2 à l'hyperplan).

Le SVM est l'un des algorithmes les plus cités dans la littérature en raison de ses performances, de sa généralité et de ses fondations théoriques [58, 59].

Les avantages du classifieur (SVM) peuvent être résumés comme suit:

1. Le classifieur (SVM) produit des classificateurs très précis lorsque les données ont une distribution inconnue.
2. Il est robuste au bruit et moins ajusté au risque.
3. Il peut résoudre tout problème complexe en spécifiant une fonction différente du noyau.

Les principaux inconvénients de l'algorithme SVM sont la difficulté de choisir la fonction du noyau appropriée et le temps de formation long pour les grands ensembles de données.

III.3.1.3 Naïve bayésienne

La classification naïve bayésienne est un type de classification Bayésienne probabiliste simple basée sur la théorie de Bayes représentant une référence théorique pour les approches statistiques de résolution des problèmes de classification.

Le principe de cette théorie est le suivant:

Soit X_j le vecteur forme dont la classe est inconnue et qu'on veut la déterminer, et soit C_i une hypothèse (X_j appartient à la classe C_i par exemple). On cherche à déterminer $P(C_i/X_j)$

la probabilité de vérification de C_i après l'observation de X_j .

$P(C_i/X_j)$ est la probabilité postérieure (après la connaissance de X_j) tandis que $P(C_i)$ est la probabilité à priori représentant la probabilité de vérification de C_i pour n'importe quel vecteur forme.

Le théorème de Bayes propose une méthode de calcul de $P(C_i/X_j)$ en utilisant les probabilités

$P(C_i)$, $P(X_j)$ et $P(X_j/C_i)$:

$$P(C_i/X_j) = \frac{P(C_i)P(X_j/C_i)}{\sum_{k=1}^C P(C_k).P(X_j/C_k)} \quad \forall i \in [1 \dots C] \quad (3.6)$$

$P(C_i/X_j)$ est donc la probabilité d'appartenance de X_j à la classe C_i , $P(C_i)$ la probabilité d'apparition de la classe C_i (peut être calculée comme le rapport entre le nombre d'échantillons appartenant à la classe C_i et le nombre total d'échantillons). $P(X_j/C_i)$ peut être considérée comme la probabilité d'apparence de chaque valeur des attributs de X_j dans les attributs des échantillons appartenant à la classe C_i . Alors l'algorithme cherche à déterminer pour chaque élément X_j la classe C_i qui maximise la probabilité de contenir cet élément.

L'avantage du classifieur bayésien naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. En effet, l'hypothèse d'indépendance des variables permet de se contenter de la variance de chacune d'entre elle pour chaque classe, sans avoir à calculer de matrice de covariance [61].

III.3.1.4 Réseau de neurone

III.3.1.4.1 Définition

Les réseaux de neurones sont une des composantes importantes du domaine de l'intelligence artificielle. Les réseaux de neurones « artificiels », pour les différencier des réseaux de neurones biologiques (cf. Tableau IV.1), sont composés de neurones artificiels simples, petites fonctions mathématiques, qui permettent, montés en réseau de former des fonctions complexes très utiles. Par analogie aux neurones biologiques, les neurones artificiels ont pour but de reproduire des raisonnements « intelligents » d'une manière artificielle. Ces neurones peuvent adopter de certaines qualités habituellement propres au biologique, c'est-à-dire, la généralisation, l'évolutivité, et une certaine forme de déduction.

Le neurone artificiel est un modèle mathématique simplifié du neurone biologique, il présente un certain nombre d'entrées, un corps traitant les entrées, et un axone véhiculant la réponse du neurone.

La Figure (Figure III.4) montre la structure d'un neurone artificiel. Chaque neurone reçoit un nombre variable d'entrées en provenance des neurones amont. A chacune de ces entrées est associé un poids W_i abréviation de Weight (poids en Anglais) représentatif de l'entrée X_0 .

Chaque neurone élémentaire est doté d'une fonction d'activation (cf. Tableau IV.2) qui donne une sortie unique y , et se ramifie ensuite pour alimenter un nombre variable de neurones avals [62].

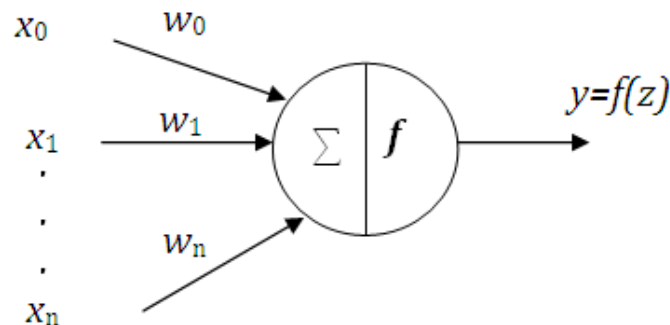


Figure III.4 Structure d'un neurone artificiel.

$$z = \sum_{i=0}^n W_i X_i \text{ et } y = f(z) \tag{3.7}$$

Σ : Fonction de combinaison, f : Fonction d'activation

Tableau III.1 Analogie entre les neurones biologiques et artificiels.

Neurone biologique	Neurone artificiel	Représentation symbolique
Soma	Neurone	
Dendrite	Intrant (<i>Input</i>)	x_i, X (intranet net)
Axone	Extrant (<i>Output</i>)	Y
Synapse	Poids	w_i

III.3.1.4.2 Naissance

En 1943, Warren McCulloch et Walter Pitts ont proposé un des premiers modèles de neurone artificiel, qui deviendra la base des réseaux de neurones artificiels. Ce modèle de neurone (également appelé neurone de McCulloch-Pitts) utilise la sommation des intrants pesés (*weighted inputs*) comme intrant « net » X (intranet direct de la fonction du neurone) et comme fonction d'activation, la fonction de signe (comparé avec le seuil θ).

$$X = \sum_{i=1}^n x_i w_i \tag{3.8}$$

$$Y = \begin{cases} -1, & x < \theta \\ +1, & x \geq \theta \end{cases} \tag{3.9}$$

Tableau III.2 Exemple de fonctions d'activation (comparées au seuil θ)

Étage	Signe	Sigmoïde	Linéaire
$Y^{étage} = \begin{cases} 1 & \text{si } X \geq \theta \\ 0 & \text{si } X < \theta \end{cases}$	$Y^{signe} = \begin{cases} +1 & \text{si } X \geq \theta \\ -1 & \text{si } X < \theta \end{cases}$	$Y^{sigmoïde} = \frac{1}{1 + e^{-\lambda(X-\theta)}}$	$Y^{linéaire} = X - \theta$

III.3.1.4.3 ARCHITECTURE

Un réseau de neurones c'est un ensemble de neurones en réseau, de sorte que les signaux sortant (*outputs*) des neurones deviennent des signaux entrant (*inputs*) dans d'autres neurones. L'architecture générale des réseaux de neurones consiste en la représentation des neurones en couches (*layers*) successives, la première représentant la couche d'intrants (*input layer*), la dernière étant la couche de sortie (*output layer*), les couches intermédiaires étant les couches cachées (*hidden layers*) du réseau. Ces couches sont dites cachées car de l'extérieur du réseau, on ne peut analyser clairement leur fonctionnement (cf. Figure III.5). On ne connaît vraiment que les signaux intrants et extrants du réseau.

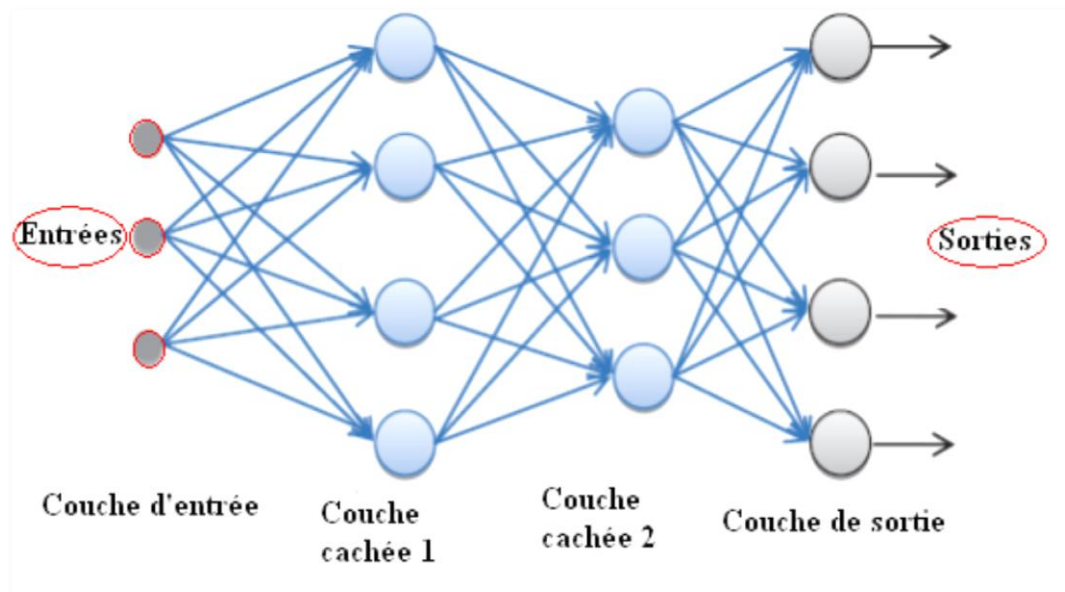


Figure III.5 Architecture d'un réseau de neurones (avec deux couches cachées).

III.3.2 Méthodes de Classification Non Supervisées

La classification non supervisée (clustering) dans laquelle le nombre de classes, inconnu a priori, est déduit directement des données. Les données disponibles ne possèdent pas d'étiquettes ; il appartient alors au système d'extraire une règle d'appartenance de chaque image à un groupe donné. Dans ce cas, les classes sont construites en fonction des données, selon l'algorithme de classification employé. L'intérêt des méthodes *non supervisées* est qu'elles ne nécessitent aucun apprentissage et par là même aucune tâche préalable d'étiquetage manuel.

III.4 Conclusion

Dans ce chapitre, nous avons présenter les notions fondamentales des différents méthodes de classification supervisée et non supervisée , en citant les plus utilisées dans l'imagerie médicale ainsi que ses principaux aspects.

nous avons intéressé à la classification supervisée des images mammographies utilisées pour le cancer du sein en particulier.

nous avons cité le rôle et l'importance des méthodes de classification dans le diagnostic et la détection des anomalies. Généralement, dans la mise en œuvre du classifieur dans, les techniques de classification supervisée

nous avons donné une vue générale sur quelques clasifieurs telque: l'algorithme "Support Vector Machine" ou "Séparateur à Vaste Marge" (SVM), Les K-plus proches voisins (KNN), Naïve Bayes et enfin les réseaux de Neurones artificiels (ANN). Nous avons cité le principe de chaque classifieur, leurs avantages et leurs inconvénients.

CHAPITRE IV

Implémentation et résultats

CHAPITRE IV Implémentation et résultats

IV.1 Introduction

La méthode de classification automatique du cancer du sein reste une stratégie difficile à résoudre, malgré les efforts scientifiques des chercheurs.

Parmi ces causes de difficultés c'est la variabilité des images mammographiques, et la variété des types de cancers qui existe, etc.

Dans ce chapitre, nous exposons l'architecture de notre système d'aide à la diagnostic automatique du cancer de sein à partir d'images mammographiques, nous présentons notre approche qui se base sur 3 étapes principales: le prétraitement, l'extraction et la sélection des caractéristiques, et la classification.

On va donc montrer, discuter et évaluer les résultats obtenus pour l'extraction des caractéristiques, ainsi que pour la classification basée sur plusieurs classifieurs SVM, NB, ANN, et KNN.

IV.2 Description des bases de données (cancer du sein)

Les bases d'images sont une ressource essentielle dans le développement des systèmes assistés par ordinateur aidé à la décision. Ils aident les chercheurs Pour mesurer les performances, à tester et à évaluer les systèmes d'aide au diagnostic. Dans le domaine du cancer du sein, il existe plusieurs bases d'images : **MIAS** (Mammographic Image Analysis Society), **DDSM** (Digital Database for Screening Mammography), **AMDI** (Indexed Atlas of Digital Mammograms) et **WDBC** (Wisconsin Diagnosis Breast Cancer).

IV.2.1 La base de donnée (MIAS)

Mammographic Image Analysis Society (**MIAS**) [75] c'est la plus ancienne base d'images publiquement disponible ; elle est encore largement utilisée dans l'état de l'art. La MIAS contient 161 cas pour un total de 322 images numérisées obtenues en incidence **MLO**, sur lesquelles sont présentés toutes les pathologies a savoir les lésions bénignes et malignes, mais également les images normales. Cette base dispose d'un nombre important de masses spicules, ainsi que d'une information relative à la densité du sein ; toutefois, la classification des masses réalisée par les radiologues ayant évalué cette base ne respecte pas les standards de l'ACR.

Dans notre thèse on a basé sur (**MIAS**), elle est choisie dans la mesure où elle est fréquemment utilisée dans la littérature. Elle permet donc de pouvoir fournir un point de comparaison plus aisé avec d'autres travaux.

IV.2.2 Informations détaillées de la base de données MIAS

Tableau IV.1 Partie de liste répertoriée des images de la base MIAS.

Numéro image	Caractère du tissu de fond	Classe d'anomalie	Gravité de l'anomalie	coordonnées du centre de l'anomalie		Rayon approximatif d'un cercle contenant l'anomalie
				x	y	
mdb001	G	CIRC	B	535	425	197
mdb063	D	MISC	B	546	463	33
mdb028	F	CIRC	M	338	314	56
mdb007	G	NORM				

La liste suivante (cf. Tableau IV.1) répertorie les films de la base de données MIAS et fournit les détails appropriés, comme suit:

1ère colonne:

Numéro de référence de la base de données MIAS.

2ème colonne:

Caractère du tissu de fond

F gras

G gras-glandulaire

D dense-glandulaire

3ème colonne:

Classe d'anomalie présente

CALC Calcification

CIRC Masses bien définies / circonscrites

SPIC Masse spiculée

MISC Autres masses mal définies

ARCH Distorsion architecturale

ASYM Asymétrie

NORM Normal

4ème colonne:

Gravité de l'anomalie:

B bénigne

M maligne

5ème, 6ème colonnes:

x, y coordonnées image du centre de l'anomalie

7ème colonne:

Rayon approximatif (en pixels) d'un cercle contenant l'anomalie.

Il y a aussi quelques points à noter:

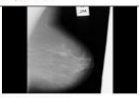
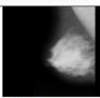

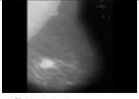
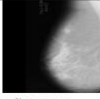

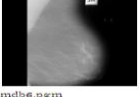
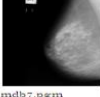

La liste est organisée en paires de films, chaque paire représentant les mammographies gauche (numéro de nom de fichier pair) et droite (numéros de nom de fichier impair) d'un même patient.

La taille de toutes les images est de 1024 pixels x 1024 pixels. Les images ont été centrées dans la matrice.

Lorsque des calcifications sont présentes, les centres et les rayons s'appliquent aux grappes plutôt qu'aux calcifications individuelles. L'origine du système de coordonnées est le coin inférieur gauche.

Dans certains cas, les calcifications sont largement réparties dans l'image plutôt que concentrées sur un seul site. Dans ces cas, les emplacements et les rayons du centre sont inappropriés et ont été omis.

IV.2.3 Echantillons de la base (MIAS)

	F (gras)	G (gras-glandulaire)	D (dense-glandulaire)
B (benignes)	 mdb012.pgm	 mdb001.pgm	 mdb005.pgm
M (malignes)	 mdb008.pgm	 mdb009.pgm	 mdb006.pgm
Normal	 mdb006.pgm	 mdb007.pgm	 mdb009.pgm

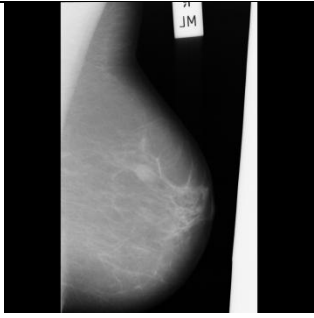
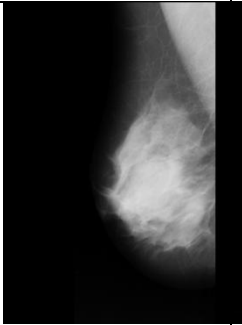
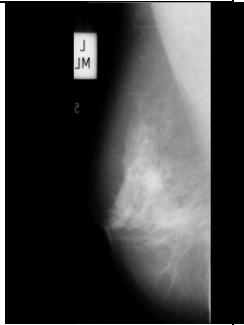
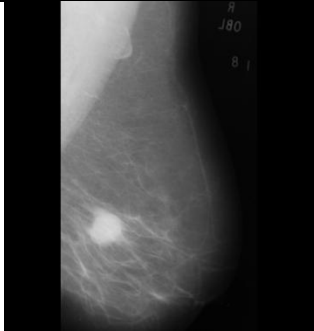
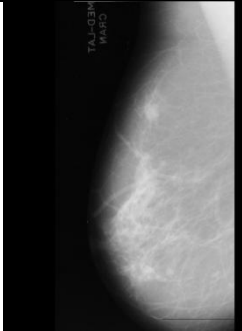
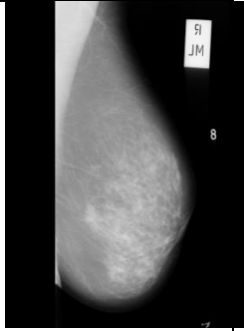
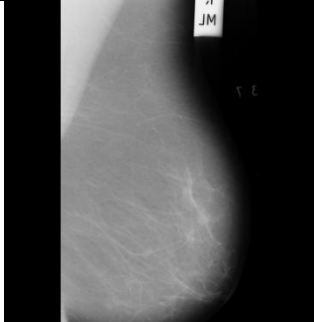
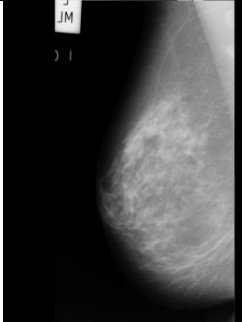
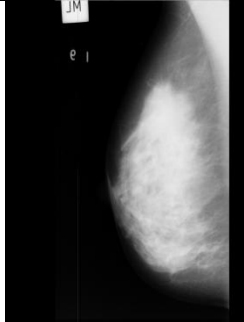
	F (gras)	G (gras-glandulaire)	D (dense-glandulaire)
B (bénigne)	 <p>Mdb12.pgm</p>	 <p>mdb01.pgm</p>	 <p>Mdb63.pgm</p>
M (malignes)	 <p>mdb28.pgm</p>	 <p>mdb23.pgm</p>	 <p>Mdb58.pgm</p>
Normal	 <p>mdb6.pgm</p>	 <p>mdb7.pgm</p>	 <p>mdb3.pgm</p>

Figure IV.1 Caractère du tissu de fond.

IV.3 le système d'aide au diagnostic (CAD) proposé

Notre système d'aide au diagnostic (CAD) proposé pour la classification des tissus mammaires dans les mammographies comme malignes ou bénignes est divisé en trois étapes principales: Le prétraitement, l'extraction et sélection des caractéristiques, et la classification (cf. Figure IV.2)

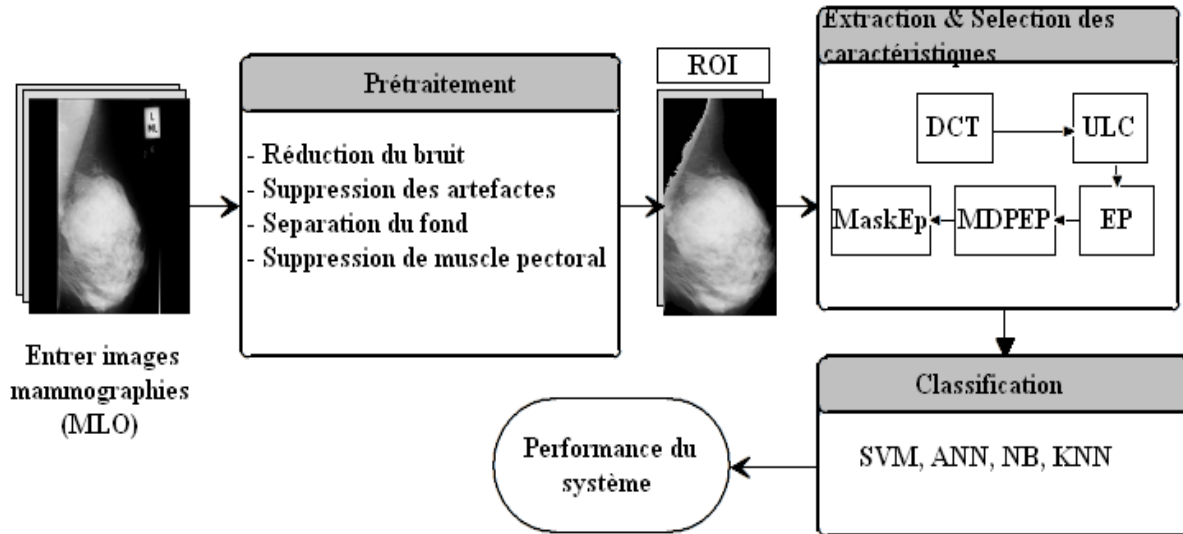


Figure IV.2 Diagramme du système proposé.

IV.3 .1 L'étape de prétraitement

L'étape de prétraitement des mammographies joue un rôle important dans la sélection de la région d'intérêt (ROI) contenant les anomalies. Cette opération consiste en diverses opérations visant à améliorer la qualité de l'image et a faciliter la segmentation. Ces opérations sont principalement la réduction du bruit, la suppression des artefacts et la séparation du fond (cf. Figure IV.4). L'utilisation des méthodes efficaces de traitement d'image représente une étape importante dans la réalisation des systèmes diagnostic (CAD) précise et très performant.

a) La réduction du bruit

Il existe différents types de bruit dans les images de mammographie, tels que le bruit de sel et de poivre, et le bruit gaussien. Traditionnellement, le filtre médian est utilisé pour éliminer ces

bruits; en raison de son comportement non linéaire, de sa simplicité et de sa capacité à préserver les arêtes [90]. Le filtre principal consiste à remplacer chaque valeur de pixel par la médiane des valeurs de pixels voisines dans la fenêtre. Dans cet travail, nous avons utilisé un filtre médian (3x3) pour réduire le bruit dans les images de mammographie (cf. Figure IV.3).

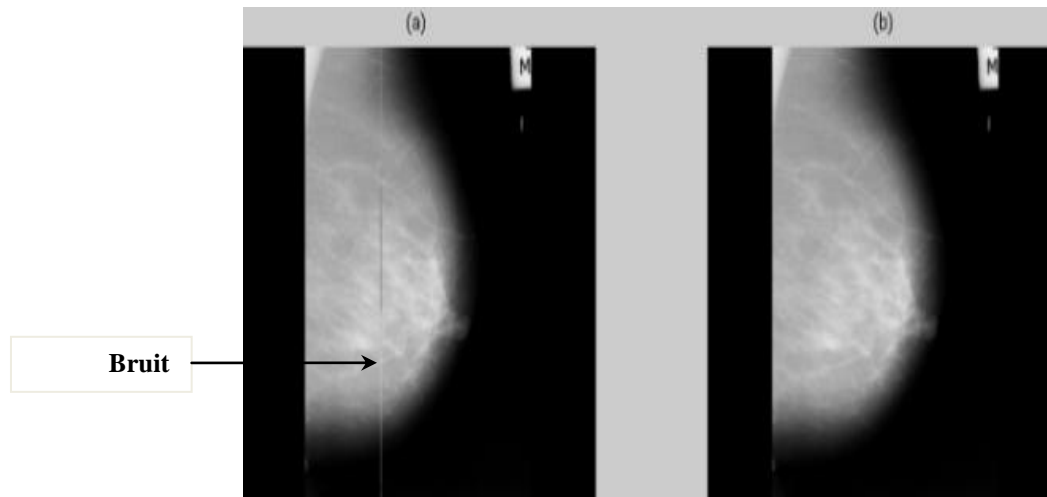


Figure IV.3 Réduction du bruit avec le filtre médian.

b) Suppression des artefacts et séparation de fond

Les étapes de l'algorithme pour supprimer tous les artefacts, marqueurs, étiquettes et séparer le profil du sein sont données comme suit [70]:

- * Binarisation de l'image en niveaux de gris avec une valeur seuil de 0,0706 [70].
- * Etiquetage de tous les objets dans l'image (c.-à-d. Artefacts, coins et étiquettes).
- * La surface de chaque objet est calculée, puis la plus grande surface est sélectionnée comme profil du sein.
- * Le résultat de l'étape précédente est utilisé comme masque de l'image originale de mammographie en niveaux de gris (cf. Figure IV.4).

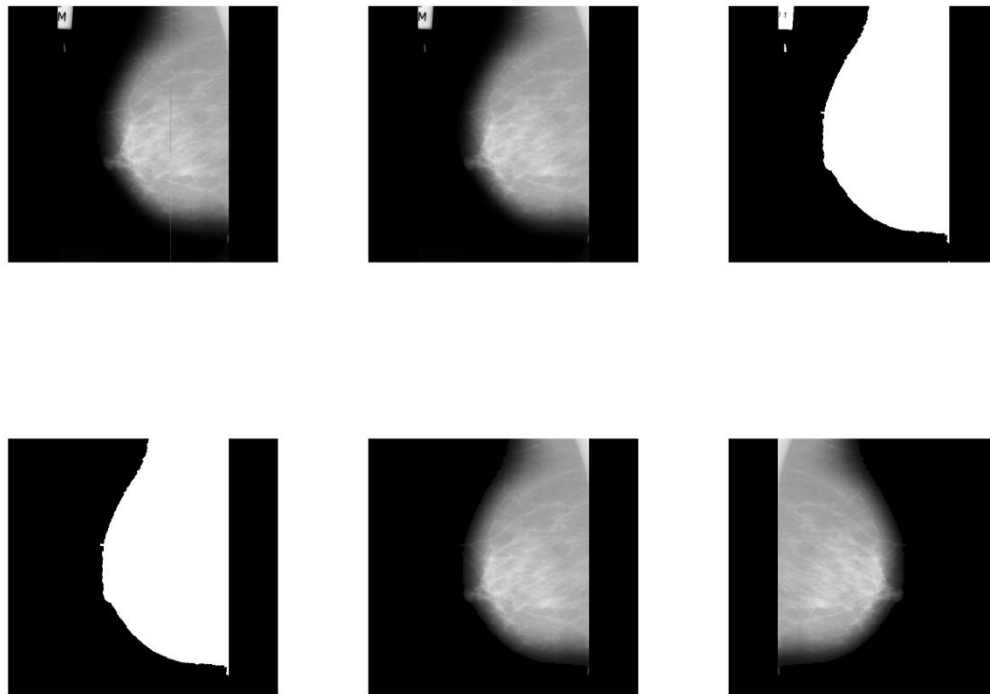


Figure IV.4 Suppression des artefacts et séparation de fond.

c) Suppression du muscle pectoral

La suppression du muscle pectoral est une tâche cruciale lors de l'étape de prétraitement dans la plupart des vues (MLO) des mammographies. Cette étape est importante pour obtenir une classification plus précise dans les systèmes de diagnostic (CAD).

Dans cette thèse, nous présentons une nouvelle approche pour la suppression de muscle pectorale basée sur deux phases:

(i) Une nouvelle méthode pour localiser la région triangulaire du muscle pectoral (**LTRPM**) est proposée, comme indiqué ci-dessous:

Algorithm 1: Localisation de la région triangulaire du muscle pectoral (LTRPM)

Entrer	MLO mammographie de taille 1024x1024
Étape (1)	Toutes les mammographies MLO droites sont retournées gauche-droite
Étape (2)	Division de MLO en quatre quadrants égaux de 512x512 pixels.
Étape (3)	Le quadrant supérieur gauche est recadré en supprimant le fond gauche.
Étape (4)	L'image résultante est divisée diagonalement en deux triangles égaux.
Sortie	Le triangle supérieur gauche ABC (contient le muscle pectoral)

(ii) Algorithme de croissance des régions (SRG)

Ensuite, nous appliquons l'algorithme (SRG) dans la région ABC obtenue (Figure IV.5) qui contient le muscle pectoral. Le point de départ est sélectionné automatiquement en tenant compte des résultats obtenus par l'algorithme 1, on prend comme ligne et comme colonne (10,10), ou (5,5). L'avantage d'appliquer la méthode (SRG) à la région triangulaire localisée est de ne supprimer que le muscle pectoral, sans supprimer complètement la région triangulaire comme dans d'autres méthodes. Les figures (IV.5), (IV.6) et (IV.7) montrent un schéma visuel de l'approche proposée.

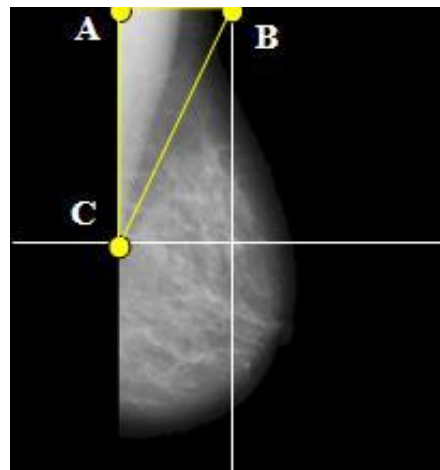


Figure IV.5 Représentation de la localisation triangulaire du muscle pectoral.

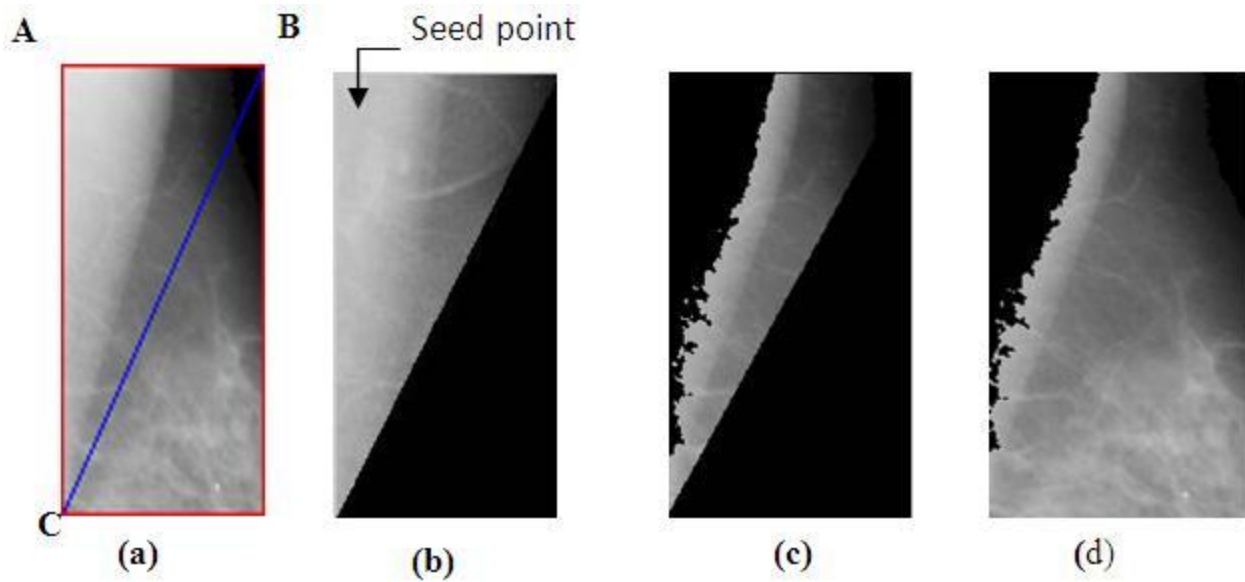


Figure IV.6 Les étapes de suppression du muscle pectoral.

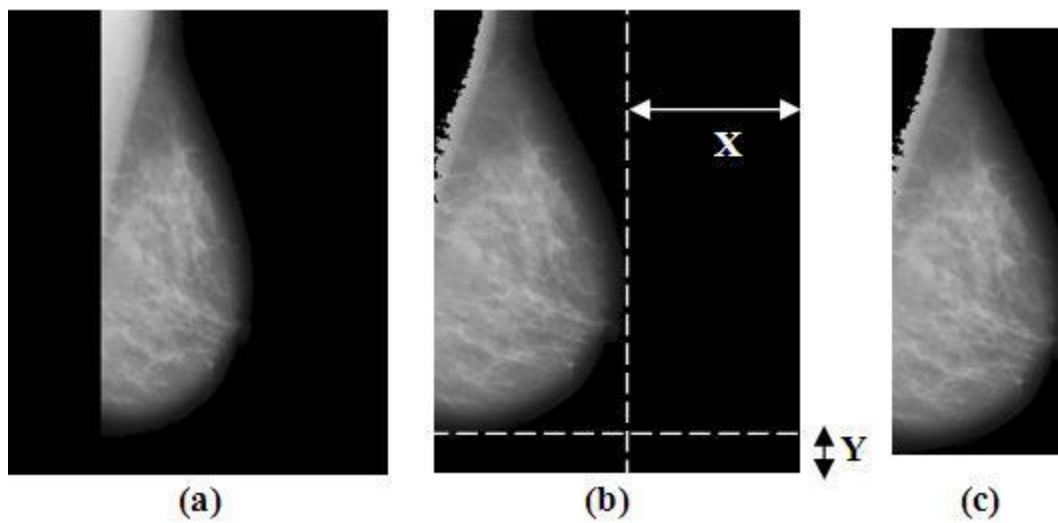


Figure IV.7 Suppression du muscle pectoral.

L'image obtenue à la figure IV.7 (b) est recadrée selon X et Y, la région d'intérêt sélectionnée (ROI) étant représentée à la figure IV.7 (c). Tous (ROIs) de chaque image de mammographie sont transformées aux mêmes dimensions.

IV.3.2 Extraction et sélection de caractéristiques

L'extraction de caractéristiques joue un important rôle et une difficile étape dans la classification précise et dans le taux élevé de diagnostic des mammographies. Dans notre travail, l'extraction des caractéristiques est appliquée dans deux domaines: le domaine fréquentiel (DCT) et le domaine d'ondelette (DWT).

a) Transformée en cosinus discrète (DCT)

Le (DCT) est une méthode mathématique, et réelle transformation pour convertir un signal ou d'image du domaine spatial au domaine fréquentiel. Il a été largement utilisé dans le traitement numérique du signal et des images, où ses principaux avantages résident dans la transformation de l'image d'entrée en coefficients réels et dans la concentration des coefficients d'informations dans la région des basses fréquences.

Il est défini par l'équation (1):

$$F(u, v) = \frac{1}{\sqrt{MN}} \alpha(u)\alpha(v) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j) \cos\left(\frac{(2i+1)u\pi}{2M}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \quad (4.1)$$

avec $0 \leq u \leq M$, $0 \leq v \leq N$, et $\alpha(u)$, $\alpha(v)$ sont défini par l'équation (4.2), (4.3)

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0 \\ 1 & \text{autre} \end{cases} \quad (4.2)$$

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{2}} & v = 0 \\ 1 & \text{otherwise} \end{cases} \quad (4.3)$$

et $F(u, v)$ est la matrice de coefficients DCT de l'image $I(i, j)$.

La sélection des coefficients DCT est une tâche fondamentale du processus d'extraction de caractéristiques, alors que ces coefficients n'ont pas la même capacité de distinguer différentes classes. Les approches classiques pour sélectionner les coefficients DCT (par exemple, les balayages en zigzag ou le masquage zonal) ne sont pas toujours efficaces pour générer les coefficients les plus discriminants.

b) Probabilité d'énergie (EP)

L'énergie est l'une des propriétés de l'image conservées dans le domaine fréquentiel après l'application de la transformation DCT. A partir de l'image transformée DCT résultante $F(u, v)$, la probabilité d'énergie de chaque coefficient DCT est définie par l'équation suivante (4):

$$EP(u, v) = \frac{|F(u, v)|^2}{\sum_{u=1}^M \sum_{v=1}^N |F(u, v)|^2} \quad (4.4)$$

La magnitude de chaque (EP) est utilisée comme critère de sélection des informations. Ainsi, les grandes valeurs d'EP représenteront la position de la plupart des coefficients DCT discriminants.

c) Transformée en ondelettes discrète (DWT)

La transformée en ondelettes (Mallat, 1989) a été proposée au milieu des années 80. Elle a été utilisée dans divers domaines tels que le traitement du signal, le traitement d'images, la vision par ordinateur, la compression d'images, la médecine biochimique, etc.

La transformée en ondelettes est une nouvelle transformation qui donne une représentation temps-fréquence d'un signal. Elle a généralement été utilisée pour la décomposition du signal (image) au moyen d'une série de filtres passe-haut afin d'analyser les composantes hautes fréquences et de filtres passe-bas pour analyser les composantes basses fréquences. Elle a été développé pour surmonter les petites variations du signal dans le temps qui ne sont pas bien couvertes par la transformée de Fourier dans le domaine fréquentiel. Mathématiquement, la transformation en ondelettes est obtenue à l'aide de l'équation suivante (4.5):

$$W_x(a, b) = \frac{1}{\sqrt{a}} \sum_{n=0}^{N-1} s(n)_N \psi^* \left(\frac{n-b}{a} \right) \quad (4.5)$$

Où a est la dilatation ou l'échelle $a = 2^j$ avec $j = 0, 1, 2, 3, \dots, \log_2(N)$, b est la translation ou l'échelle $b = a2^j$ et $s(n)$ est le bruit et ψ est l'ondelette mère. Dans les approches DWT, les résultats des filtres passe-haut et passe-bas sont appelés les coefficients DWT; coefficients d'approximation $A(k)$ et coefficients de détail $D(k)$, ces coefficients DWT permettent la reconstruction du signal à l'aide de l'équation (4.6), et les filtres d'analyse et de synthèse sont identiques, à l'exception du retournement temporel.

$$s(n) = \sum_{m=1}^L \left[\sum_{k=-\infty}^{\infty} D_m(k) \psi_{m,k}(t) + \sum_{k=-\infty}^{\infty} A_l(k) \phi_{l,k}(t) \right] \quad (4.6).$$

IV.3.2.1 Première partie (domaine fréquentiel DCT)

Dans ce travail, une nouvelle méthode d'analyse de la puissance des plus coefficients discriminantes (MDPEP) pour l'extraction des caractéristiques est proposée. Il se compose de trois phases. Tout d'abord, la transformée en cosinus discrète (DCT) est appliquée sur toutes les régions d'intérêt obtenues (ROIs), puis les coefficients DCT du coin supérieur gauche (ULC) sont sélectionnés. Dans la deuxième phase, la méthode de probabilité d'énergie utilisée pour la reconnaissance de visage dans [74] est utilisée comme critère de sélection des informations effectives appliquées sur la région ULC. Dans la dernière phase, une combinaison de la méthode EP et de l'algorithme des coefficients de puissance discriminants [63] est proposée pour calculer la matrice de puissance de discrimination MDP (i, j). Cette matrice est définie comme le rapport entre la variance interclasse et la variance intra-classe. Par conséquent, des valeurs élevées signifient une précision de classification élevée. L'algorithme 2, ci-dessous montre comment calculer la matrice des plus forts discriminants caractéristiques (MDP_{ij}). Considérant une image A_{ij} de taille $N \times M$. D_{ij} sont les coefficients transformés par la 2D-DCT et E_{ij} est la probabilité d'énergie des D_{ij} . La base de données contient C classes et pour chaque classe il y'a S images. $C \times S$ images sont présentées pour l'apprentissage.

Algorithme 2: Algorithme des plus puissants coefficients Discriminatives (MDPEP)

En entrée: Nombre de classes C , Nombre images d'apprentissage S pour chaque classe

$Im=zeros(N, M)$ // matrice zéro initial

Requiert: fonction *FoncPT* () de Prétraitement, la transformée en cosinus discrète 2 dimension DCT2.

```

(1) For chaque classe c  $\forall c \leq C$ 
(2)      $S_{ij} = Im;$ 
(3)     For chaque image s  $(A_{ij}(S, C)) \quad \forall s \leq S$ 
(4)          $A_{ij}(S, C) \leftarrow \text{FoncPT}(A_{ij}(S,C));$  // appel FoncPT()
(5)          $D_{ij}(S, C) \leftarrow \text{DCT2}(A_{ij}(S,C));$ 
           // Choisir ULC (coefficients DCT)
(7)          $E_{ij}(S, C) = (\text{abs}(D_{ij}(S,C)))^2 / \text{sum}(\text{sum}(D_{ij}(S,C).^2));$ 
(8)          $S_{ij} \leftarrow S_{ij} + E_{ij}(S,C)$  // la somme des images pour chaque classe
(9)     End
(10)     $M_{ij}^c \leftarrow S_{ij} / S;$  // la moyenne pour chaque classe
(11)     $V_{ij}^c = Im;$ 
(12)    For chaque image s  $(A_{ij}(S,C)) \quad \forall s \leq S$ 
(13)         $A_{ij}(S,C) \leftarrow \text{FoncPT}(A_{ij}(S,C));$ 
(14)         $D_{ij}(S,C) \leftarrow \text{DCT2}(A_{ij}(S,C));$ 
           // Choisir ULC (coefficients DCT)
(15)         $E_{ij}(S, C) = (\text{abs}(D_{ij}(S,C)))^2 / \text{sum}(\text{sum}(D_{ij}(S,C).^2));$ 
(16)         $V_{ij}^c \leftarrow V_{ij}^c + (E_{ij}(S,C) - M_{ij}^c).^2;$  // la variance pour chaque classe .
(17)    End
(18) End
(19)     $M_{ij} \leftarrow \frac{1}{C} \sum_{c=1}^C M_{ij}^c;$  // la moyenne des images de toutes les classes.
(20)     $V_{ij}^w \leftarrow \frac{1}{C} \sum_{c=1}^C V_{ij}^c;$  // la variance intra classe.
(21)     $V_{ij}^B = Im;$ 
(22) For chaque classe c  $\forall c \leq C$ 
(23)     For chaque image s  $(A_{ij}(S, C)) \quad \forall s \leq S$ 
(24)          $A_{ij}(S, C) \leftarrow \text{FoncPT}(A_{ij}(S, C));$ 
(25)          $D_{ij}(S, C) \leftarrow \text{DCT2}(A_{ij}(S,C));$ 
           // Choisir ULC (coefficients DCT)
(26)          $E_{ij}(S, C) = (\text{abs}(D_{ij}(S, C)))^2 / \text{sum}(\text{sum}(D_{ij}(S,C).^2));$ 
(27)          $V_{ij}^B \leftarrow V_{ij}^B + (E_{ij}(S,C) - M_{ij})^2;$  //la variance inter classe
(28)     End
(29) End
(30)     $MDP(i,j) = V_{ij}^B / V_{ij}^w;$  //  $1 \leq i \leq M, 1 \leq j \leq N$ 

```

Sortie $MDP(i,j)$ // la matrice de la puissant discrimination

Les grandes valeurs de MDP impliquent un grand pouvoir de discrimination des coefficients. Les coefficients (Eij) sont sélectionnés de manière adaptative en fonction de leurs valeurs MDP correspondantes, où les coefficients (Eij) avec un MDP supérieur sont conservés [63]. Par conséquent, un masque (MaskEp) est généré comme indiqué à la Figure IV.8 La matrice MDP est convertie en un vecteur de colonne, qui est trié par valeurs dans un ordre décroissant.

Le nombre des caractéristiques sélectionnées dépend du seuil appliqué aux valeurs MDP triées. Afin de sélectionner **k** caractéristiques, nous choisissons une valeur seuil égale à la valeur MDP à la position **k**. Ensuite, la matrice de masque est créée en sélectionnant la matrice MDP avec la valeur de seuil sélectionnée.

Le masque contient les positions (valeurs égales à 1) des caractéristiques les plus discriminantes. Lors de l'étape de classification, le masque binaire créé (**MaskEp**) sera utilisé pour générer les vecteurs de caractéristiques.

$$MaskEp(i, j) = \begin{cases} 1, & MDP(i, j) \geq Th(k) \\ 0, & \text{autre.} \end{cases} \quad (4.7)$$

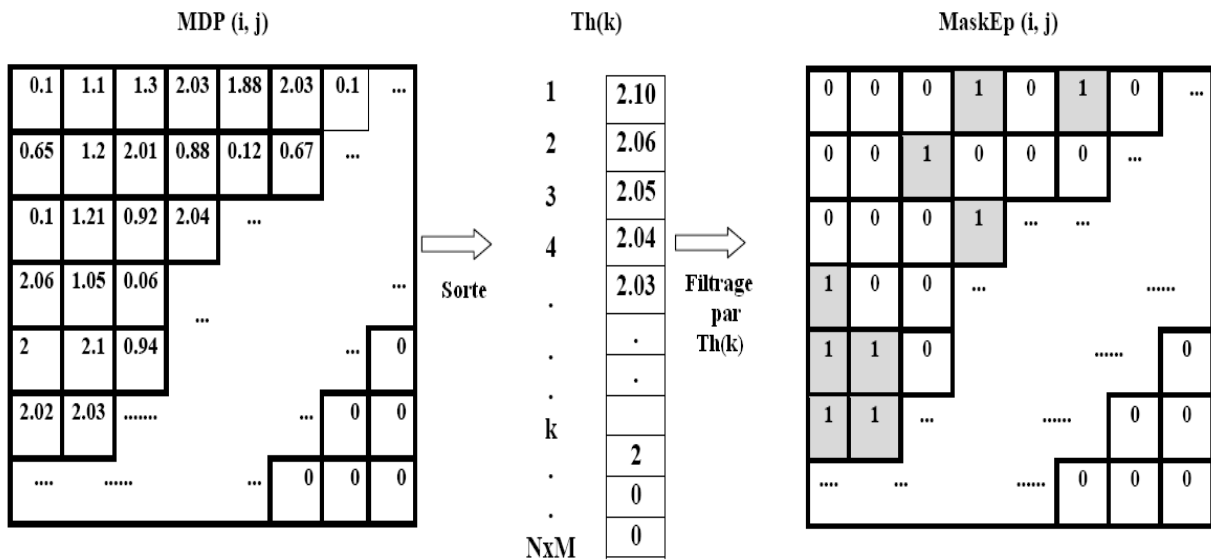


Figure IV.8 Création du masque binaire.

IV.3.2.2 deuxième Partie (domaine ondelettes DWT)

Notre système d'aide au diagnostic (CAD) proposé pour la classification des tissus mammaires dans les mammographies comme malignes ou bénignes dans le domaine ondelette est illustré dans le digramme en dessous (cf. Figure IV.9).

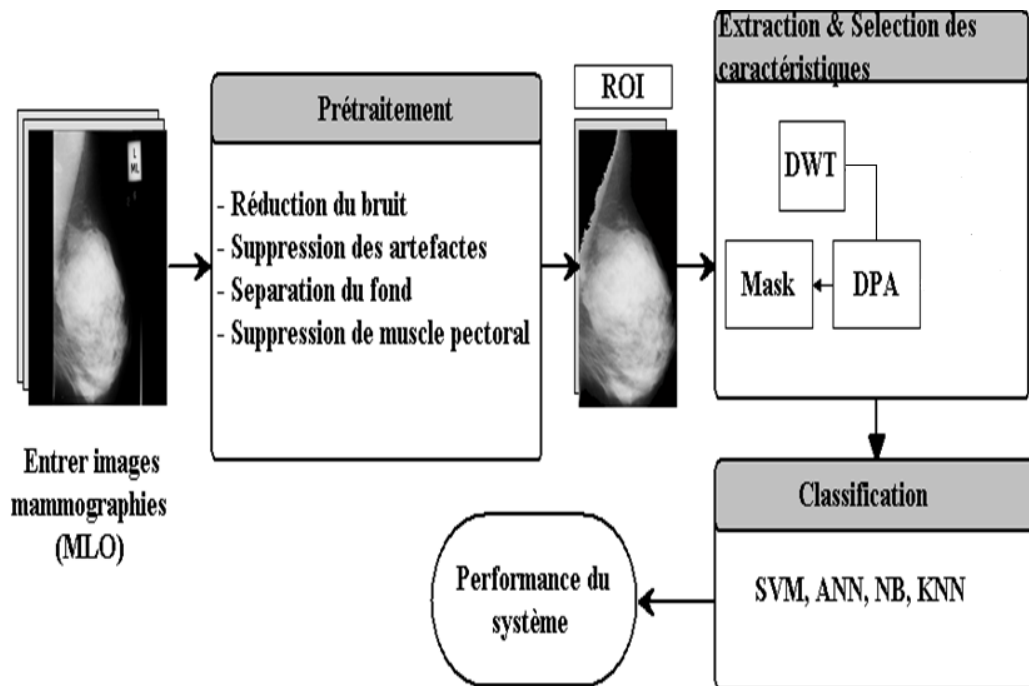


Figure IV.9 Diagramme du système proposé dans le domaine en ondelette.

On calcule le (MDP_{ij}) dans le domaine transformé en ondelette. Considérant une image A_{ij} de taille $N \times M$. E_{ij} sont les coefficients calculés par la fonction $dwting(n, A_{ij})$ voir figure (Figure IV.10), avec n : niveau de décomposition. La base de données est la même comme précédemment.

- dans cette partie, la création du masque binaire (Mask) suite les mêmes étapes comme précédent

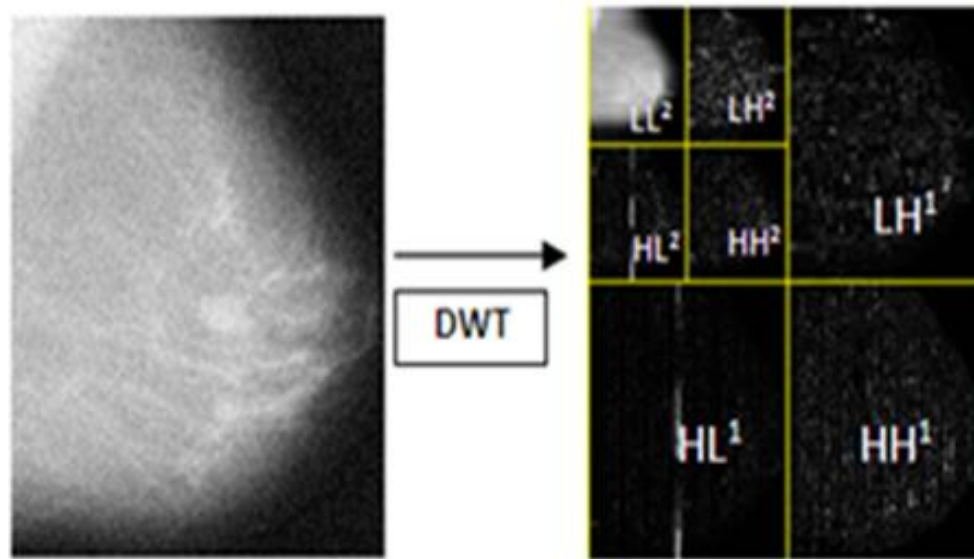


Figure IV.10 décomposition en 2 niveaux avec `dwtimg()` [95].

IV.3.3 Classification

La classification des images joue un rôle important dans le domaine du diagnostic médical. Par conséquent, les images doivent être classées avec une précision maximale. En conséquence, plusieurs méthodes de classification automatisées ont été proposées. Dans cette section, nous avons évalué quatre types différents de classifieurs SVM, ANN, NB et KNN afin de classer la tumeur du sein comme bénigne ou maligne.

a) Machine à vecteurs de support (SVM)

Le SVM un classifieur linéaire binaire non probabiliste. Pratiquement, il passe par deux phases: une phase d'apprentissage et une phase de test. la phase d'apprentissage construit un modèle qui permettra de classer les futures données inconnues. Dans la phase de test on va utiliser la structure créée précédemment pour classer tes échantillons inconnus, et en fin on peut calculer la précision du classifieur.

b) Réseau de neurones artificiels (ANN)

Les réseaux de neurones artificiels sont des réseaux à réaction qui peuvent être formés pour classer les entrées en fonction des classes cibles. Le réseau de neurones contient une architecture à trois couches: la couche d'entrée, la couche cachée et la couche de sortie [38]. Dans notre cas, la couche d'entrée est basée sur le nombre des caractéristiques sélectionnées, la couche cachée contient 10 neurones et enfin la couche sort qui diagnostique si elle est bénigne ou maligne.

c) classificateur Naïve Bayes (NB)

Naïve Bayes est un algorithme d'apprentissage automatique supervisé basé sur l'application du théorème de Bayes, basé sur des probabilités conditionnelles. Un classifieur Naïve Bayes attribue une nouvelle observation à la classe la plus probable, en supposant que les entités sont conditionnellement indépendantes pour une valeur donnée de la classe. Il est facile et rapide de prédire la classe de l'ensemble de données de test, mais le principal inconvénient de ce dernier est son exigence de prédictors indépendants.

d) k-voisins les plus proches (KNN)

On peut schématiser le fonctionnement de K-NN en l'écrivant en pseudo-code suivant :

$D = \{(x', c), c \in C\}$ les données d'apprentissage.

Soit x l'ensemble dont on veut déterminer leur classe

Algorithme 3: KNN

En entrée: un ensemble d'apprentissage D avec $D = \{(x', c), c \in C\}$.

Une fonction de définition distance d .

Un nombre entier k

Début

pour chaque $((x', c) \in D)$ **faire**

 calculer la distance euclidienne $d(x, x')$

Fin

pour chaque $= \{x' \in K_{ppv}(x)\}$ **faire**

 compter le nombre d'occurrence de chaque classe

Fin

Attribuer à x la classe la plus fréquente;

Fin

IV.4 Résultats et discussions

Pour valider le système proposé, des expériences ont été effectuées sur la base de données MIAS (Mammographic Image Analysis Society) [75]. Nous avons testé les performances de la méthode proposée en utilisant le langage MATLAB en calculant des mesures métriques telles que la précision, la sensibilité et la spécificité à l'aide des expressions suivantes:

$$\text{Sensibilité} = TP / (TP + FN) \times 100 (\%) \quad (4.8)$$

$$\text{Spécificité} = TN / (TN + FP) \times 100 (\%) \quad (4.9)$$

$$\text{Précision (Accuracie)} = (TN + TP) / (TN + FN + FP + TP) \times 100 (\%) \quad (4.10)$$

Où:

TP (vrai positif) est le nombre des malins classés comme malins.

TN (vrai Négative) est le nombre des bénins classés comme bénins.

FP (Faux Positive) est le nombre des malins classés comme bénins.

FN (Faux Négative) est le nombre des bénins classé comme malins.

IV.4.1 Résultats du première partie

Dans cette partie, nous avons construit la matrice MDP en utilisant 50 mammographies bénignes et 50 mammographies malignes choisis au hasard en fonction de l'algorithme MDPEP. Le même nombre d'échantillons doit être sélectionné dans chaque classe. Ensuite, la base de données est divisée aléatoirement en deux parties un pour l'apprentissage et l'autre pour le test.

Nous avons pris 113 (ROIs) divisés en 70% pour l'apprentissage et 30% pour les tests. Les résultats donnés sont la moyenne de dix essais. Le Tableau IV.2 montre une comparaison entre les performances mesurées des classifieurs ANN, Naïve Bayes, SVM et KNN. On constate que la précision moyenne de la technique proposée peut atteindre 100% avec le classifieur (ANN), 94.11 avec (SVM), 92,64% avec (NB) et 87.35 avec KNN.

Tableau IV.2 Les résultats des performances de classification obtenus en utilisant les quatre classifieurs.

Classifieur	Précision (%)	Sensitivité (%)	Spécificité (%)
ANN	100	100	100
Naïve Bayes (NB)	92.64	89.47	96.66
SVM	94.11	96.23	92.17
KNN	87.35	91.05	82.67

Nous avons évalué les performances de classification de l'algorithme proposé en fonction du nombre de caractéristiques utilisées et des différentes tailles (ULC) dans la classification. Le Tableau IV.3 montre la précision, la sensibilité et la spécificité de la classification à l'aide des classificateurs SVM, ANN et NB.

Dans le Tableau IV.3, nous pouvons voir que les meilleures performances sont obtenues avec 100 caractéristiques et avec une taille de 512x512 ULC.

Tableau IV.3 performances pour différents tailles (ULC) et différents nombres de caractéristique [94].

		SVM			ANN			NB		
ULC	N. de caractéristiques	précision (%)	Sensitivité (%)	Spécificité (%)	précision (%)	Sensitivité (%)	Spécificité (%)	précision (%)	Sensitivité (%)	Spécificité (%)
1024x1024	20	80.29	86.11	75.83	90.3	90	90.5	85.29	82.63	88.66
	40	85	87.15	84.54	93.8	88	98.4	87.35	83.15	92.66
	60	88.82	95.48	83.12	98.2	98	98.4	90	84.73	96.66
	80	87.64	91.63	84.86	96.5	100	93.7	90.88	89.48	92.66
	100	92.35	97.83	87.34	98.2	96	100	92.35	91.05	94
512x512	20	83.52	87.67	80.13	92	96	88.9	84.7	76.84	94.66
	40	87.35	91.99	84.33	93.8	94	93.7	88.52	82.63	96
	60	88.23	93.19	83.41	96.5	94	98.4	89.70	89.73	96
	80	90	94.04	87.2	98.2	98	98.4	91.17	88.94	97.33
	100	94.11	96.23	92.17	100	100	100	92.64	89.47	96.66
256x256	20	82.17	89.36	76.18	93.9	92	95.2	81.17	76.84	86.66
	40	80.56	85.48	78.32	92	92	92.10	82.5	74.21	92
	60	81.17	86.64	76.79	95	92	98.4	87.4	81.15	99.33
	80	82.63	86.7	78.85	95.6	96	95.2	87.94	82.10	95.3
	100	85.88	89.74	82.55	95.6	92	98.4	88.52	80.52	98.66
128x128	20	77.35	84.80	72.26	83.5	78	87.3	75.88	63.15	92
	40	77.35	80.76	74.76	86.7	84	88.9	77.5	66.31	90.66
	60	79.11	83.04	75.56	93.8	94	93.7	8.29	71.57	91.33
	80	78.52	82.41	75.41	594.	98	93.7	81.17	70	95.33
	100	81.47	85.51	77.84	94.7	94	95.2	82.05	72.10	94.66

D'autre part, la Figure IV.10 représente la précision de la classification par rapport différentes tailles d'ULC avec 100 caractéristiques.

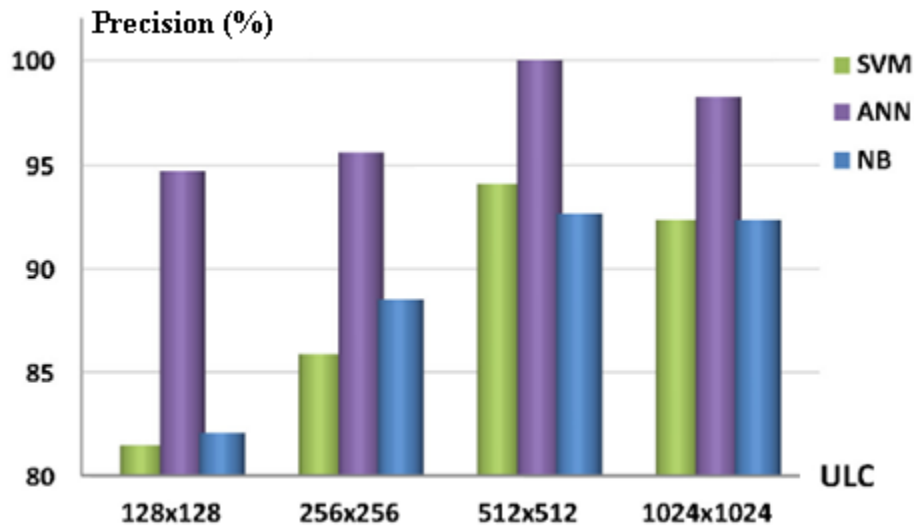


Figure IV.11 la précision de classification par rapport à la taille ULC (avec 100 caractères).

La figure IV.11 représente la variation de la précision de la classification en fonction du nombre de caractères avec une taille ULC fixe de 512x512. La figure IV.13 illustre les performances de classification de l'ANN en utilisant la matrice de confusion pour les données d'apprentissage, de test et de validation.

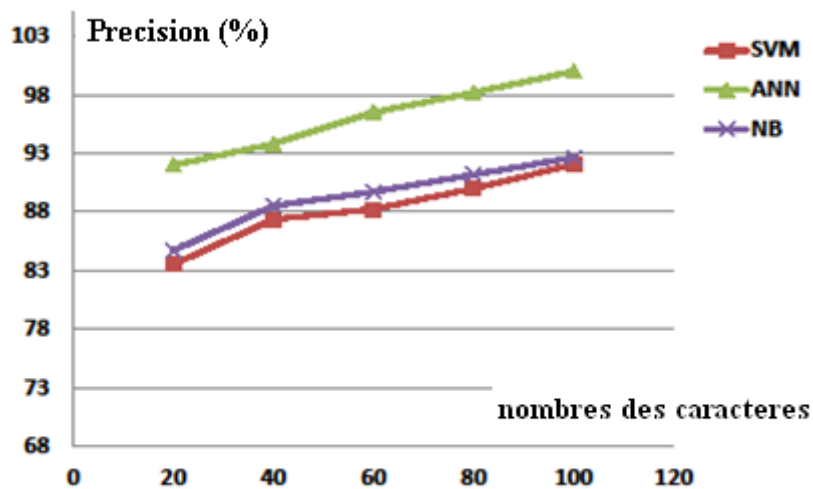


Figure IV.12 précision de la classification en fonction du nombre de caractères.

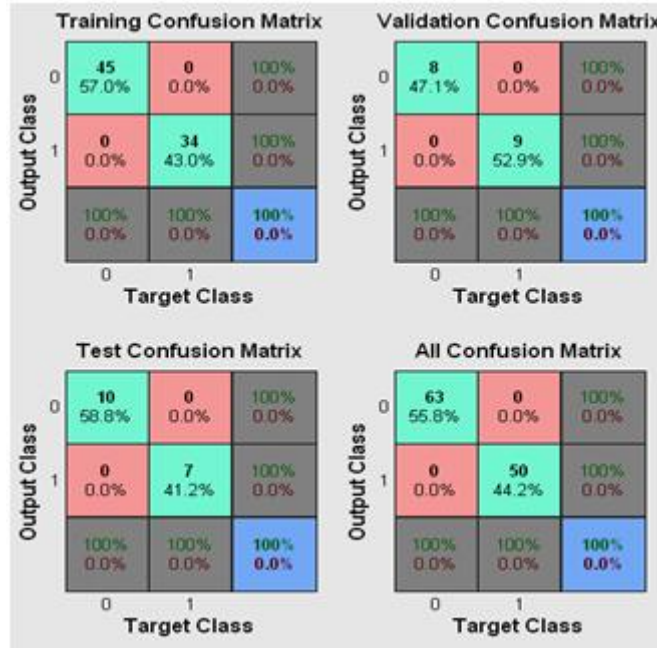


Figure IV.13 performance de classification du ANN[94].

Pour le classifieur KNN, nous avons obtenu les résultats reportés dans le Tableau IV.4 avec une taille de retour sur investissement de 512x512 et différentes valeurs de voisins (K). Les résultats sont obtenus avec le même nombre de caractéristiques 100. Nous pouvons voir que les performances du classifieur KNN augmentent avec le nombre K. Les meilleurs résultats sont obtenus avec K = 9 avec une précision de 87,35%.

Tableau IV.4 Performances de classification de KNN [94].

Taille (ROI)	512x512			
K	3	5	7	9
Précision (%)	85	85	85.8	87.35
Sensitivité (%)	86.3	86.3	90.5	91.05
Spécificité (%)	83.3	83.3	80	82.67

Les résultats obtenus dans cette session sont reportés dans le Tableau IV.5, avec une taille de 512x512 ULC et 100 caractéristiques, et on prend le même nombre (malins-bénins) pour création de la matrice MDP.

On observe que les performances de toutes les classifieurs utilisés augmentent avec l'augmentation du nombre d'images dans les jeux de données (malignes-bénignes). Il est également rapporté que les meilleurs résultats obtenus sont de 50-50.

Tableau IV.5 Performances de classification en fonction du nombre d'images (malins, bénins) [94].

<i>Précision (%)</i>			
100 caractéristiques et 512x512 ULC			
Nombres de (malins, bénins)	SVM	ANN	NB
20-20	70.88	78.8	67.35
30-30	78.52	83.2	77.64
40-40	86.17	95.6	85
50-50	94.11	100	92.64

IV.4.2 Résultat du deuxième Partie (domaine DWT)

Dans cette deuxième partie, nous avons construit la matrice MDP en utilisant 50 mammographies bénignes et 50 mammographies malignes choisis au hasard en fonction de l'algorithme 3 (DPA-DWT). Le même nombre d'échantillons doit être sélectionné dans chaque classe. Ensuite, la base de données est divisée aléatoirement en deux parties un pour l'apprentissage et l'autre pour le test.

Nous avons pris 113 (ROIs) divisés en 70% pour l'apprentissage et 30% pour les tests. Les résultats donnés sont la moyenne de dix essais.

Le Tableau IV.6 montre une comparaison entre les performances mesurées des classifieurs KNN, ANN, Naïve Bayes et SVM. On constate que la précision moyenne de la technique proposée peut atteindre **99.1%** avec le classifieur (ANN), **98,53%** avec le classifieur (NB), **99.12** avec le classifieur (SVM) et **68.94** avec le classifieur (1-NN).

Tableau IV.6 Les résultats des performances de classification obtenus.

ROI 512x512, Niveau=3, 'bior4.4'				
	Sensitivité (%)	Spécificité (%)	Précision (%)	Nombre de caractéristiques
ANN	100.00	98.00	99.1	100
Naïve Bayes	97.89	99.33	98.53	200
SVM	100	98.13	99.12	350
KNN (k=1)	75.52	60.6	68.94	100

Les Tableaux IV.7, IV.8, IV.9 représentent les variations de la précision, de la Sensitivité et la Spécificité en fonction du nombre de caractéristiques avec une taille (ROI) fixe de 512x512 et de 3 niveaux de décomposition avec l'ondelette 'Bior4.4' pour SVM, ANN, NB respectivement.

Tableau IV.7 Les résultats des performances de classification pour SVM.

SVM ROI (512x512), Niveau=3, 'bior4.4'			
Nombres de caractéristiques	Sensitivité (%)	Spécificité (%)	Précision (%)
10	92.66	85.58	88.79
40	95.94	94.99	95.29
60	97.31	93.63	95.59
90	96.84	94.21	95.6
100	98.47	93.77	96.18
150	99.00	96.88	97.94
200	98.97	96.91	97.94
250	99.50	97.50	98.53
300	100	97.57	98.82
350	100	98.13	99.12

Tableau IV.8 Les résultats des performances de classification pour ANN.

ANN - ROI (512x512), 'bior4.4', Niveau=3			
Nombres de caractéristiques	Sensitivité (%)	Spécificité (%)	Précision (%)
10	89.4	88	90.5
40	100	96.8	98.2
60	98.4	98	98.2
80	98	95.2	96.5
90	98.4	98	98.2
100	100	98.00	99.1
150	96.8	100	98.2
200	96.8	100	98.23
250	98.4	100	99.1

Tableau IV.9 Les résultats des performances de classification pour NB.

Naïve Bayes (NB) - ROI (512x512, 'bior4.4'), Niveau =3			
Nombres de Caractéristiques	Sensitivité (%)	Spécificité (%)	Précision (%)
10	91.79	85.52	88.47
40	93.68	98.00	95.59
60	93.68	98.00	95.59
80	96.32	95.33	95.88
90	94.74	98.00	96.18
100	96.84	97.33	97.06
150	95.79	98.00	96.76
200	97.89	99.33	98.53
250	96.32	100.00	97.94
300	96.32	100.00	97.94

Pour montrer l'efficacité de la méthode proposée, nous avons comparé d'autres travaux. Le Tableau IV.10 montre une comparaison entre les résultats de notre algorithme proposé et les résultats précédents rapportés dans la littérature. Nous pouvons voir que le système de CAD proposé donne des résultats plus précis que ceux obtenus avec d'autres méthodes.

Tableau IV.10 Comparaison de Résultats avec les méthodes existantes.

Autheurs	Année	Database	Classifieur	Classes	Précision (%)
Ramani [19]	2014	MIAS	Neural network	2	89.7
			Naïve Bayes		81.21
			Random forest		86.67
Pak [30]	2015	MIAS	AdaBoost	2	91.43
Beura [31]	2015	MIAS	BPNN	2	98
Abirami [32]	2016	MIAS	Neural network (RBF)	2	93
Kumari [33]	2016	MIAS	SVM	2	76
			ANN	2	88
Singh [39]	2016	MIAS	Random Forest	2	93.9
Nusantara [35]	2016	MIAS	KNN	2	96.8
Elmoufidi [34]	2017	MIAS	SVM	2	94.4
Singh [40]	2017	MIAS	Random Forest	2	97.3
Sheba [36]	2018	MIAS	FFANN	3	96.0
			ANN		100
Méthode Proposé DCT-MDPEP	2019	MIAS	SVM	2	94.11
			Naïve Bayes		92.64
DWT-DPA	2019	MIAS	ANN	2	99.1
			SVM		99.12
			Naïve Bayes		98.53

IV.6 Conclusion

Dans ce chapitre, un système de CAD est développé pour classer les images de mammographie comme bénignes ou malignes. Le système proposé comporte plusieurs étapes. Premièrement, le prétraitement qui a été appliqué pour limiter la région de recherche d'anomalies en supprimant le bruit de fond et tous les objets indésirables, dans lequel une nouvelle méthode de suppression des muscles pectoraux est proposée. Après cela, la plupart des caractéristiques discriminantes sont extraites à l'aide de notre algorithme proposé (MDPEP) et du masque de probabilité d'énergie (MaskEp) dans le domaine fréquentiel. Dans le domaine en ondelette les caractéristiques discriminantes sont extraites à l'aide d'un algorithme qui fusionne entre la méthode d'analyse power discriminant (DPA) et DWT.

Enfin, les caractéristiques extraites sont fournies en entrée à différents classifieurs ANN, NB, KNN et SVM. La comparaison des résultats obtenus avec ceux obtenus en utilisant les techniques récemment proposées montre la supériorité de l'algorithme proposé par rapport aux autres méthodes.

Conclusion générale et perspectives

Conclusion générale et perspectives

Le cancer du sein reste le type le plus meurtrier, il est devenu un grave problème de santé publique dans de nombreux pays. Actuellement, la mammographie est utilisée pour le diagnostic précoce du cancer du sein. D'autre part, la croissance rapide des mammographies a créé un besoin pour développer des systèmes de diagnostic assisté par ordinateur (CAD) aidant les radiologues.

Dans notre travail de thèse, un système d'aide à la décision CAD a été développé pour classer les images de mammographie comme bénignes ou malignes. Le système proposé comporte plusieurs étapes. Tout d'abord, un prétraitement a été appliqué pour réduire la région de recherche contenant l'anomalie. Dans la deuxième étape, une approche hybride pour la suppression du muscle pectoral a été proposée pour sélectionner la région d'intérêt (ROI). Ensuite, les caractéristiques dans le domaine transformé (DCT et DWT) les plus discriminantes sont extraites puis sélectionnées à l'aide de l'algorithme DPA proposé. Enfin, les primitives extraites sont utilisées comme entrées pour les classificateurs KNN, ANN, NB et SVM.

Par l'application de la méthode proposée sur les images de la base de données MIAS, pour la première partie (DCT+ULC+EP+DPA) nous avons pu atteindre une précision moyenne de 94.11% avec le SVM. Par ailleurs, le classifieur KNN a atteint 87,35% de précision. Avec le classifieur ANN, nous pouvons atteindre la valeur de 100%. Avec le classifieur NB, nous pouvons atteindre une précision 92.64%.

Pour la deuxième partie (DWT+DPA) nous avons pu atteindre une précision moyenne de 99.12% avec le **SVM**. Par ailleurs, le classifieur **ANN** a atteint 99.1% de précision. Avec le classifieur **NB**, nous pouvons atteindre la valeur de 98.53%.

La comparaison entre les résultats obtenus par la méthode proposée et d'autres techniques récemment proposées montre la supériorité de notre algorithme par rapport à d'autres méthodes.

Dans nos perspectives de travail, nous nous intéressons toujours au domaine de la classification des images médicales, et spécialement les points suivants :

- ❖ L'étude de l'efficacité des caractéristiques extraites par d'autres transformations d'images appliquées aux régions d'intérêt.
- ❖ L'utilisation d'autres méthodes de segmentation d'images afin d'obtenir une suppression plus performante du muscle pectorale, et ainsi obtenir une meilleure région d'intérêt (ROI).
- ❖ L'application puis l'évaluation des caractéristiques utilisées par d'autres classifieurs plus sophistiqués tels que le deep learning.
- ❖ Elaborer une méthode de sélection optimale des caractéristiques discriminantes en fonction de plusieurs paramètres concernant le nombre de caractéristique et le classifieur lui-même.

Liste des productions scientifiques

▪ International Publication

Benhassine NE, Boukaache A, Boudjehem D. Classification of mammogram images using the energy probability infrequency domain and most discriminative power coefficients. *Int J Imaging Syst Technol.* 2019;1-12. <https://doi.org/10.1002/ima.22352>.

Catégorie A , Impact factor 1.423

▪ International Communications

1- **Nasser Edinne Benhassine**, Mohamed Rouis, Abdelnour Boukaache, Djalil Boudjehem and Abdelkrim Ouafi, " A Comparative Study of Image De-noising based Wavelet Thresholding Techniques In mammography images", Second International Conference on Electrical Engineering ICEEB'2018, December 2-3 2018, Biskra, Algeria.

2- Mohamed Rouis, **Nasser Edinne Benhassine**, Abdelkrim Ouafi, Salim Sbaa, Abdelnour Boukaache and Djalil Boudjehem, "Noise Removal by the Discrete Wavelet Transform in Electrocardiography signal (ECG)", Second International Conference on Electrical Engineering ICEEB'2018, December 2-3 2018, Biskra, Algeria.

Bibliographie

- [1] Richard Zimmermann. "La médecine nucléaire:La radioactivité au service du diagnostic et de la thérapie". EDP Sciences, 2006.
- [2] Cheikhrouho Imene.Description et classification des masses mammaires pour le diagnostic du cancer du sein,thèse pour l'obtention du titre de Docteur en Traitement des images et du signal,27 juin 2012.
- [3] S.Rethice théophile. "Contribution a l'analyse d'images médicales pour la reconnaissance du cancer du sein", thèse , université paris-saclay, 2016.
- [4] Smith, R. A. (2003). Iarc handbooks of cancer prevention, volume 7: Breast cancer Screening. Breast Cancer Research, 5(4):216–217.
- [5] Mechta Chahinase.interpretation des images demammographie par l'ontologie,mémoire pour l'obtention du diplôme de magister en informatique,2012/2013.
- [6] ACR, of Radiology, A. C., et al. (2003). Breast imaging reporting and data system atlas (Bi-rads atlas). Reston, VA : American College of Radiology, 98. (Cite en pages 19 et 21.)
- [7] Chekkaf Hamza, Touil Ibrahim.Segmentation des masses dans les images mammographiques,mémoire pour l'obtention du diplôme de Master en Informatique,20112012.
- [8] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., and Sickles., E. A. (2003). American college of radiology (breast imaging reporting and data system). SFR (Société Française de Radiologie).
- [9] Eccles, S. A., Aboagye, E. O., Ali, S., Anderson, A. S., Armes, J., Berditchevski,F., Blaydes, J. P., Brennan, K., Brown, N. J., Bryant, H. E., et al. (2013). Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Research, 15(5):1. (Cite en page 20.).
- [10] Mammographic density and the risk and detection of breast cancer, *NEJM* , 2007,356:227-36, Boyd NF et al. "Mammographic density and the risk and detection of breast cancer". *NEJM N Engl J Med* 2007 Jan 18 ; 356:227-36.
- [11] H.fanet,Imagerie médicale à base de photons: Radiologie, tomographie X, tomographie gamma et positons, imagerie optique. Lavoisier, 2010.
- [12] M. Wallis, M. Walsh and J. Lee, A review of false negative mammography in a ymptomatic population, Clin. Radiology, vol.44, pp.13-15, 1991.

- [13] M. L. Giger, H.-P. Chan, and J. Boone. Anniversary paper : History and status of CAD and quantitative image analysis : the role of Medical Physics and AAPM. *Medical physics*, 35(12):5799–5820, 2008.
- [14] Suzuki K. A review of computer-aided diagnosis in thoracic and colonic imaging. *Quantitative Imag Med Surg*. 2012;2(3):163–76.
- [15] H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis : an ROC study. *Radiology*, 212(3) :817–27, 1999.
- [16] J. C. Dean and C. C. Ilvento. Improved Cancer Detection Using Computer-Aided Detection with Diagnostic and Screening Mammography : Prospective Study of 104 Cancers. *American Journal of Roentgenology*, 187(1) :20–28, 2006.
- [17] Zakeri, Fahimeh Sadat, Hamid Behnam, and Nasrin Ahmadinejad. "Classification of benign and malignant breast masses based on shape and texture features in sonography images." *Journal of medical systems* 36.3: 1621-1627, 2012.
- [18] Lahmiri, Salim, and Mounir Boukadoum. "Hybrid discrete wavelet transform and Gabor filter banks processing for features extraction from biomedical images." *Journal of medical engineering*, 2013.
- [19] Campos, Lucio FA, Aristófanés C. Silva, and Allan Kardec Barros. "Diagnosis of breast cancer in digital mammograms using independent component analysis and neural networks." *Iberoamerican Congress on Pattern Recognition*. Springer, Berlin, Heidelberg, 2005.
- [20] Martins, Leonardo de O., et al. "Classification of normal, benign and malignant tissue using co-occurrence matrix and Bayesian neural network in mammographic images." *Neural Networks, 2006. SBRN'06. Ninth Brazilian Symposium on*. IEEE, 2006.
- [21] Bala, B. Kiran, and S. Audithan. "Wavelet and curvelet analysis for the classification of microcalcification using mammogram images." *Current Trends in Engineering and Technology (ICCTET)*, 2014 2nd International Conference on. IEEE, 2014.
- [22] H. Chan et al., Computerized analysis of mammographic micro calcifications in morphological and feature spaces, *Medical Physics*, vol.25, no.10, pp.2007-2019, 1998.
- [23] S. Lai, X. Li and W. Bischof, On techniques for detecting circumscribed masses in mammograms, *IEEE Trans. Medical Imaging*, vol.8, no.4, pp.377-386, 1989.

- [24] D. Brazokovic and M. Neskovic, Mammogram screening using multiresolution-based image segmentation, *International Journal of Pattern Recognition and Artificial Intelligence*, vol.7, no.6, pp.1437- 1460, 1993.
- [25] H. Li et al., Markov random field for tumor detection in digital mammography, *IEEE Trans. Medical Imaging*, vol.14, no.3, pp.565-576, 1995.
- [26] I. Christoyianni et al., Fast detection of masses in computer-aided mammography, *IEEE Signal Processing Magazine*, pp.54-64, 2000.
- [27] M. A. Jaffar, A. Hussain, A. M. Mirza and A. Chaudhry, Fuzzy entropy and morphology based fully automated segmentation of lungs from CT scan images, *International Journal of Innovative Computing, Information and Control*, vol.5, no.12(B), pp.4993-5002, 2009.
- [28] L. F. A. Campos, A. C. Silva and A. K. Barros, Diagnosis of breast cancer in digital mammograms using independent component analysis and neural networks, *X Iberoamerican Conference on Pattern Recognition, LNCS*, vol.3773, pp.460-469, 2005.
- [29] D. D. Costa, A. K. Barros and A. C. Silva, Independent component analysis in breast tissues mammograms images classification using LDA and SVM, *Information Technology Applications in Biomedicine, Conference on the 6th International Special Topic, Tokyo*, pp.231-234, 2007.
- [30] Priyanka kamboj and versha rani –A Brief study of various noise model and filtering techniques versha rani et al, *journal of global research in computer science*, 4 (4), april 2013, 166-171.
- [31] Harish Kumar.N,Amutha.S and Dr.Ramesh Babu D.R. Enhancement of Mammographic Images using Morphology and Wavelet Transform.
- [32] Er. Manpreet Kaur , Er. Gagndeeep Kaur A P. A Survey on Implementation of Discrete Wavelet Transform for Image Denoising, *International Journal of Communications Networking System*.
- [33] Andreas RICK,« Représentation de la variabilité dans le traitement d’images flou Application à la mammographie numérique ». Thèse présentée pour l’obtention du titre Docteur de l’université paris 6.
- [34] J.P Coquerez. S. P. Foliguet, "*Analyse d'image: filtrage et segmentation*," Edition Masson, 1995.
- [35] Hamza CHEKKAF, Ibrahim TOUIL «Segmentation des masses dans les images mammographiques ». Thèse présentée pour l’obtention du titre de Docteur. Université Abou Bakr Belkaid Tlemcen, *le 25 septembre 2012*.
- [36] Cyril Meurie.2005"Segmentation d’images couleur par classification pixellaire et hiérarchie de partitions" doctorat de l’université de CAEN.

- [37] M.Meliani 2012"Segmentation d'Image par Coopération Régions-Contours". Thèse de magister,Ecole nationale Supérieure en Informatique, Oued-SmarAlger, Ecole Doctorale STIC.
- [38] M. Sandeli. 2014 " Traitement d'images par des approches bio-inspirées". Thèse de magister,Université Constantine 2.
- [39] Jalalian A,et al. Review article : Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection, *EXCLI journal*, (2017) 16:113-137.
- [40] El Atlas N, El Aroussi M, Wahbi M. Computer-aided breast cancer detection using mammograms: A review. Paper presented at the Second World Conference on Complex Systems (WCCS), 2014.
- [41] J. K. KIM et H.W. PARK (1999). Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3):231–238.
- [42] Z. HUO, M. L.GIGER et al (1998). Automated computerized classification of malignant and benign masses on digitized mammograms. *Academic Radiology*, 5(3):155–168.
- [43] Filipczuk P, Fevens T, Krzyżak A, Obuchowicz A. GLCM and GLRLM based texture features for Computer aided breast cancer diagnosis. *J Med Inform Technol*. 2012;19:109-15.
- [44] M. J. Islam, M. Ahmadi, and M. A. Sid-Ahmed, "An Efficient Automatic Mass Classification Method in Digitized Mammograms Using Artificial Neural Network," *International Journal of Artificial Intelligence - Applications (IJAIA)*, vol. 1, no. 3,1-13, 2010.
- [45] Liu GH, Yang JY. Content-based image retrieval using color difference histogram. *Pattern Recogn*. 2013;46:188-98.
- [46] Shen L, Rangayyan RM, Desautels JL. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*. 1994;13(2):263-74.
- [47] M. KUPINSKI, et M. GIGER, (1999). Feature selection with limited datasets. *Medical Physics*, 26(10): 2176–2182.
- [48] M. KUPINSKI, et M. GIGER, (1997). Feature selection and classifiers for the computerized detection of mass lesions in digital mammography. *International Conference on Neural Networks*,4:2460–2463.
- [49] Michie, D., Spiegelhalter, D., & C.C. (1994). *Machine learning, neural and statistical classification*. New York: Ellis Horwood.
- [50] Mari, J., & Napoli, A. (1996). *Aspects de la classification*. Rapport technique 2909, INRIA.

- [51] Henriet, L. (2000). Système d'évaluation et de classification multicritères pour l'aide à la décision, construction de modèles et procédures d'affectation. *Thèse de doctorat en science*. Université Paris Dauphine.
- [52] C. LURETTE, " *Développement d'une technique neuronale auto-adaptative pour la classification dynamique de données évolutives. Application à la supervision d'une presse hydraulique* ", Thèse de Doctorat, Université des sciences et technologies de LILLE 2003.
- [53] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern classification*. John Wiley & Sons, 2001.
- [54] M. Goldstein, "k-Nearest Neighbor Classification," *Information Theory, IEEE Transactions on*, vol. 18, pp. 627-630, 1972.
- [55] Blum, A. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *In Artificial Intelligence*.
- [56] Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer Verlag.
- [57] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3) :273–297.
- [58] Shawe-Taylor, J. and Cristianini, N. *Kernel methods for pattern analysis*. (2004).Cambridge Univ.
- [59] Rakotomamonjy, A. (2003). Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, pages 3 :1357–1370.
- [60] Rajpoot, K. M., & Rajpoot, N. (2004). Wavelets and support vector machines for texture classification. *In: Proceedings of 8th International Multitopic Conference*, 328–333.
- [61] Ouarda ASSAS, "Classification floue des images", THESE DOCTORAT EN SCIENCES, Université de Batna, soutenu publiquement le 14/12/2013
- [62] R. GHAYOULA, " *Contribution à l'Optimisation de la Synthèse des Antennes Intelligentes par les Réseaux de Neurones* ", Thèse de Doctorat, Laboratoire de Physique de la Matière Molle, Unité de Recherche des Circuits et Systèmes d'Electroniques à Haute Fréquence, Tunisie, 2008.
- [63] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology." *Pattern recognition*. **43**(4), 1431-1440 (2010).
- [64] V.P. Singh, R. Srivastava, "Content-Based Mammogram Retrieval Using Wavelet based Complete LBP and K-means Clustering for the Diagnosis of Breast Cancer," *International Journal of Hybrid Intelligent System*, **14**, 31-39 (2017)

- [65] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Physics in medicine & biology*. **43**(2), 365 (1998).
- [66] F. Georgsson, and M. Yaffe, "Transformation of mammograms based on anatomical features," *Proc. Int. Workshop of Digital Mammography, ed. M. Yaffe, Medical Physics Publishing*. (2000).
- [67] S. M. Kwok, R. Chandrasekhar, and Y. Attikiouzel, "Automatic pectoral muscle segmentation on mammograms by straight line estimation and cliff detection," *Intelligent Information Systems Conference, the Seventh Australian and New Zealand 2001. IEEE*, (2001).
- [68] W.Xu, S. Xia, and H. Duan, "A novel computer-aided diagnosis system of the mammograms," *Intelligent Computing in Signal Processing and Pattern Recognition. Springer, Berlin, Heidelberg*. pp. 639-644 (2006).
- [69] R. Ramani, and N. S. Vanitha, "Computer Aided Detection of Es in Mammograms," *International Journal of Image, Graphics and Signal Processing*, **6**(4), 54 (2014).
- [70] J. Nagi et al., "Automated breast profile segmentation for ROI detection using digital mammograms," *Biomedical Engineering and Sciences (IECBES)*. (2010).
- [71] A. O. Malagelada, "Automatic Mass Segmentation in Mammographic Images," *PHD Thesis, University of Girona, Girona*, April, (2007).
- [72] V.P. Singh, R.Srivastava, "Automated and effective content-based mammogram retrieval using wavelet based CS-LBP feature and self-organizing map," *Biocybern Biomed Eng*, (2017).
- [73] V. P. Singh, S. Srivastavab, and R. Srivastavaa, "An Efficient and Automated Content Based Image Retrieval for Digital Mammography," *Journal of X-Ray Science and Technology*, **26** pp. 29-49 (2018),
- [74] C. Jean, C. Yun-Su, K. Ki-Hyun, and Y. Jang-Hee, "Face Recognition Using Energy Probability in DCT domain," *ICME-IEEE*, pp. 1549–1552 (2006) [doi: 10.1109/ICME.2006.262839].
- [75] J Suckling *et al* (1994). *The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series*, 1069 pp375-378(1994).
- [76] R. Adams, and L. Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*. **16**(6), 641-647 (1994).
- [77] C. Wang, X. Mai, G. Lin, and C.Tan Kuo, "Classification for Breast MRI Using Support Vector Machine," *International Conferences on Computer and Information Technology Workshops, IEEE*, pp 362- 367 (2008).
- [78] V. Vapnik, "Statistical Learning Theory", Wiley. (1998).
- [79] The World Health Organization (WHO):
<http://www.who.int/mediacentre/factsheets/fs297/en/>

- [80] Ganesan K, Acharya UR, Chua CK, Min LC, Abraham KT, Ng KH. Computer-aided breast cancer detection using mammograms:A review. *IEEE Reviews in Biomedical Engineering*. 6: 77-98. 2013.
- [81] Srivastava S, Sharma N, Singh SK, Srivastava R. Quantitative analysis of a general framework of a CAD tool for breast cancer detection from mammograms. *Journal of Medical Imaging and Health Informatics*. 4(5): 654-74. 2014.
- [82] S. Sapate, and S. Talbar, "An Overview of Pectoral Muscle Extraction Algorithms Applied to Digital Mammograms," *Medical Imaging in Clinical Applications*. Springer, pp. 19-54 (2016).
- [83] F. Pak, H. R. Kanan, and A. Alikhassi. "Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution," *Computer methods and programs in biomedicine*, **122**(2), 89-107 (2015).
- [84] S. Beura, B. Majhi, and R. Dash. "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer." *Neurocomputing*, **154**, 1-14 (2015).
- [85] C. Abirami, H. Rajaguru, and S. Chakravarthy, "Performance analysis and detection of micro calcification in digital mammograms using wavelet features," *Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on. IEEE*, (2016).
- [86] V. Kumari, K. U. Rani "Mammogram classification using multilayer perceptron and support vector machine," *International Journal of Computational Science, Mathematics and Engineering*. 3(1), 2349-8439 (2016).
- [87] A. Elmoufidi et al., "Automatic detection of suspicious lesions in digital X-ray mammograms," *Advances in Ubiquitous Networking 2. Springer, Singapore*. pp. 375-385 (2017).
- [88] A. C. Nusantara, E. Purwanti, and S. Soelistiono, "Classification of digital mammogram based on nearest neighbor method for breast cancer detection," *International Journal of Technology*. **1**(1), 71-77 (2016).
- [89] K. U. Sheba, and S. G. Raj, "An approach for automatic lesion detection in mammograms," *Cogent Engineering*. **5**(1), 1444320 (2018).
- [90] Kshema, J. George M, and D. A.S. Dhas, "Preprocessing Filters for Mammogram Images: A Review," *IEEE Conference on Emerging Devices and Smart Systems*, pp. 3-4 (2017).
- [91] S. Singh, P. R. Gupta, "Breast cancer detection and classification using neural network," *International Journal of Advanced Engineering Sciences and Technologies*. **6**, 4-9 (2011).

- [92] V. P. Singh et al., " Mammogram Classification Using Selected GLCM Features and Random Forest Classifier," *International Journal of Computer Science and Information Security*, **14**(6), 82 (2016).
- [93] V. P. Singh, S. Srivastava, and R. Srivastava, "Effective Mammogram Classification Based on Center Symmetric-LBP features in Wavelet Domain Using Random Forest Classifier," *Technology and Health Care*. **1**,1-19 (2017).
- [94] Benhassine NE, Boukaache A, Boudjehem D. Classification of mammogram images using the energy probability infrequency domain and most discriminative power coefficients. *Int J Imaging Syst Technol*. 2019;1-12. <https://doi.org/10.1002/ima.22352>.
- [95] Benhassine NE, M. Rouis, A. Boukaache, D. Boudjehem et A. Ouafi, " A Comparative Study of Image De-noising based Wavelet Thresholding Techniques In mammography images", Second International Conference on Electrical Engineering ICEEB'2018, December 2-3 2018, Biskra, Algeria.