

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université Saad Dahlab-Blida1

Faculté des Sciences

Département d'Informatique

Année Universitaire 2021-2022



Projet de fin d'études
pour l'obtention du diplôme de Master en Informatique

Spécialité : Système Informatique et Réseau (SIR)

Thème :
Conception d'un système d'aide au diagnostic
médical

Réalisé par :

GOITA EMMANUEL

Soutenu le 27 septembre 2022 devant le jury :

Mr. M. HAMMOUDA

Président

Dr. H. DAOUD

Examineur

Dr. S. BOUDRAA

Encadreur

Dédicace

A ma chère mère,

A mon cher père,

A mes frères et sœurs,

A mes oncles,

A toute ma famille de près ou de loin,

A mes amis,

A mes collègues.

EMMANUEL

Remerciements

*Je remercie **Dieu** de m'avoir accordé des connaissances de la science et de nous avoir aidé à réaliser ce travail.*

Au terme de ce modeste travail je tiens à remercier chaleureusement et respectivement tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste projet de fin d'étude.

*Nos vifs remerciements vont tous d'abord à notre encadreur **Mme. Boudraa Sawsen**. J'exprime ma profonde gratitude de m'avoir encadré tout le long de ce projet et qui m'a apporté beaucoup de connaissances dans ce vaste domaine de « système d'aide au diagnostic médical ».*

Tous mes respects et mes remerciements vont vers les membres du jury qui vont pleinement consacrer leur temps et leur attention afin d'évaluer ce travail, qui j'espère le sera à la hauteur de leur attente.

Enfin, mes remerciements les plus sincères sont adressés à tous les professeurs, l'administration et le personnel de l'informatique qui ont contribué à forger nos connaissances et à assister notre formation.

Résumé :

Le cancer du sein étant l'un des cancers les plus mortels, il est un réel danger et soucis dans la santé car il cause le décès de plusieurs femmes atteintes par cette maladie. Vue sa complexité et sa difficulté d'être détecté à l'état précoce cela donne l'avantage au tumeur de se développer au fur du temps sans connaissance, ce qui amène vraisemblablement une mort certaine de la victime car faute de le découvrir avant son développement dans le sein. C'est pour cela nous avons jugé mettre en place un système de diagnostic assisté par ordinateur qui analyse automatiquement les mammographies numériques et donner un second avis aux médecins pour un traitement efficace, cela enlève plusieurs charges sur le médecin et aide à travailler plus efficacement et à sauver la vie des patientes atteintes. Ce système de diagnostic aurait besoin d'un ensemble d'images mammographiques provenant de la base MIAS à son entrée , il aurait plusieurs étapes qui sont les suivantes : segmentation des masses (tumeurs), application de la méthode des motifs locaux binaires, l'extraction des caractéristiques statistiques du premier ordre et la classification. Cette classification est la finalité du système, à travers des méthodes de classification, le système départage les images qui ont un tumeur maligne d'avec celles qui ont tumeur bénigne.

Mots clés : Système d'aide au diagnostic, Classification, Cancer du sein, Image mammographique, extraction caractéristiques, segmentation k-means.

Abstract :

Breast cancer is one of the most deadly cancers and is a real health danger as it causes the death of many women with this disease. The complexity of the disease and the difficulty of detecting it early gives the tumour the advantage of growing unnoticed over time, which is likely to lead to the victim's certain death. Therefore, we have decided to set up a computer aided diagnosis system which automatically analyses digital mammograms and gives a second opinion to the doctors for effective treatment, which removes many charges on the doctor and helps to work more efficiently and save the lives of the affected patients. This diagnostic system would need a set of mammographic images from MIAS as input, it would have several steps which are : segmentation of masses (tumors), application of local binary pattern method, extraction of first order statistics features and classification. This classification is the purpose of the system, through classification methods, the system separates the images that have a malignant tumour from those that have a benign tumour.

Keywords : Diagnosis aid system, Classification, Breast cancer, Mammographic image, Feature extraction, K-means segmentation.

ملخص :

يعتبر سرطان الثدي من أكثر السرطانات فتكا ، فهو خطر حقيقي وقلق صحي لأنه يتسبب في وفاة العديد من النساء يعتبر سرطان الثدي من أكثر السرطانات فتكا، فهو خطر حقيقي على الصحة لأنه يتسبب في وفاة العديد من النساء المصابات بهذا المرض. نظراً لتعقيده وصعوبة اكتشافه في مرحلة مبكرة ، فإن هذا يمنح الورم ميزة التطور بمرور الوقت دون وعي ، مما يؤدي على الأرجح إلى وفاة الضحية بشكل مؤكد. لهذا السبب قررنا تنفيذ نظام مساعدة التشخيص بالكمبيوتر، والذي يقوم تلقائياً بتحليل صور الثدي الشعاعية الرقمية وإعطاء رأي ثانٍ للأطباء للعلاج الفعال ، وهذا يزيل العديد من الأعباء على الطبيب ويساعد على العمل بكفاءة أكبر وإنقاذ حياة المرضى المصابين. سيحتاج نظام التشخيص هذا إلى مجموعة من الصور الشعاعية للثدي من قاعدة بيانات MIAS عند إدخالها ، وسيكون لها عدة مراحل وهي كالتالي: تجزئة الكتل (الأورام) ، تطبيق طريقة الأنماط المحلية الثنائية ، استخراج الخصائص الإحصائية والتصنيف. هذا التصنيف هو الغرض من النظام ، من خلال طرق التصنيف ، يقوم النظام بفصل الصور التي بها ورم خبيث عن تلك التي بها ورم حميد. الكلمات المفتاحية: نظام مساعدة التشخيص ، التصنيف ، سرطان الثدي ، صورة الثدي الشعاعية ، استخلاص الخصائص ، تجزئة k-mean.

Table des matières

Dédicace	2
Remerciements	3
Résumé :	4
Abstract :	5
ملخص.....	6
Table des matières	7
Table des figures	10
Liste des Tableaux.....	12
Liste des Abréviations	1
Introduction Générale	1
1.Contexte et problématique	1
2. Objectifs	2
3. Plan de travail	2
Chapitre 1 : Contexte médical.....	3
1.1 Introduction	3
1.2 Le cancer du sein	3
Généralités	3
1.2.1 Description du sein	4
1.2.2 Concept cancer	5
1.2.3 Définition du cancer du sein	5
1.2.3.1 Cancer fréquent	6
1.2.3.2 Les différents types de cancer du sein	7
1.2.4 Quelques symptômes du cancer du sein	8
1.2.5 Les causes d'un cancer du sein	8
1.3 Le diagnostic du cancer du sein	10
1.3.1 Comment découvre-t-on un cancer du sein	10
1.3.2 Le dépistage	11
1.3.3 Fiabilité du diagnostic de cancer du sein	11
1.3.4 Les examens complémentaires	13
1.3.4.1 Les examens radiologiques	13
1.3.4.2 Les prélèvements	14

1.3.4.3 Les examens sanguins	14
1.4 Les différents traitements du cancer du sein	15
1.5 La surveillance après le traitement	19
1.6 Vivre pendant et après un cancer	19
1.7 Conclusion	20
Chapitre 2 : Les systèmes d'aide au diagnostic médical	21
2.1 Introduction	21
2.2 Description des systèmes de Détection/Diagnostic Assistés par Ordinateur (DAO)	21
2.2.1 Importance et nécessité des systèmes DAO	21
2.2.2 Détection Assistée par Ordinateur (DAOe)	22
2.2.3 Diagnostic Assisté par Ordinateur (DAOx)	22
2.3 Composition des systèmes DAOx	23
2.3.1 Prétraitement	24
2.3.2 Segmentation	26
2.3.2.1 Les méthodes de seuillage	27
2.3.2.2 Les méthodes basées régions	27
2.3.2.3 Les méthodes basées contours	28
2.3.3 Extraction des caractéristiques : Description	29
2.3.3.1 La couleur	30
2.3.3.2 La texture	30
2.3.3.3 La forme	38
2.3.4 Classification	40
2.3.4.1 Les méthodes de classifications non supervisées	41
2.3.4.2 Les méthodes de classification supervisées	42
2.5 Quelques travaux sur les systèmes d'aide au diagnostic dédiés aux images mammographiques	43
2.6 Conclusion	46
Chapitre 3 : Segmentation via l'algorithme k-means	48
3.1 Introduction	48
Généralités	48
3.1.1 le Clustering	48
3.1.2 L'algorithme k-means	49
3.2 Méthode de partition du clustering en global	50
3.2.1 Processus de clustering K-means	50

3.3 Utilisation de l'algorithme k-means dans la segmentation des masses mammaires	52
3.3.1 Description	52
3.3.2 Matériels et méthodes utilisés dans cette segmentation de la masse	52
3.4 Les méthodes de clustering les plus utilisées	54
3.4.1 Fuzzy c-means algorithm	54
3.4.2 Méthodes basées sur la distribution	54
3.4.3 Méthodes basées sur le centroïde	55
3.4.4 Méthodes basées sur la connectivité	56
3.4.5 Modèles de densité	56
3.4.6 Agrégation de sous-espace	57
3.6 Conclusion	58
Chapitre 4 : Conception et Implémentation du système.....	59
4.1 Introduction	59
4.2 Conception du Système	59
4.2.1 Environnement et outils de travail	59
4.2.2 Images Mammographiques (MIAS)	60
4.2.3 Métriques d'évaluation	60
4.2.4 L'Organigramme du système	62
4.2.5 La segmentation par la méthode k-means	62
4.2.5.1 Expériences et résultats	63
4.2.6 Extraction des caractéristiques	64
4.2.7 Classification	64
4.3 Implémentation du système	65
Partie 1 : Segmentation	65
4.3.1 Interface	65
4.3.2 Sélection de l'image	66
4.3.3 Segmentation image	67
Partie 2 : LBP, Extraction de caractéristiques et classification.....	69
Cette partie comporte deux phases : l'apprentissage et la classification.....	69
4.4 Conclusion	82
Conclusions Générales et perspectives	83
Références Bibliographiques	84

Table des figures

Figure 1.1 Anatomie du sein [17].....	4
Figure 1.2 Anatomie du sein féminin [18]	5
Figure 1.3 Cellules cancéreuses (cellules normales, cellules cancéreuses, cancer in situ, cancer infiltrant (Rupture de la membrane, propagation par les capillaires et vaisseaux lymphatiques)) [19]	6
Figure 1.4 Cancer canalaire infiltrant et cancer canalaire in situ [20]	8
Figure 1.5 Les quatre secteurs du sein [21].....	12
Figure 1.6 Chirurgie conservatrice [22]	16
Figure 1.7 : Récepteurs hormonaux d'un cancer hormonosensible [23].....	18
Figure 2.1 – Diagramme général d'un système DAOx [39]	24
Figure.2.2. Les différents systèmes CAD (computer-aided diagnosis/Detection) [59]	29
Figure 2.5: Les voisins d'une distance $d = 2$ du pixel 'x' selon les 4 directions.	33
Figure 2.6 : exemple d'une image.....	33
Figure 2.7: La matrice de cooccurrence $M_{i,jd}$ de l'image (Fig. 2.6).....	34
Figure 2.8 : Calcul LBP sur patch 3×3 sur image	38
Figure 2.9–Schéma détaillant les notions de : fermeture, intérieur, contour et complémentarité [39].	40
Figure 3.1 : Steps for performing clustering analysis [78].....	49
Figure.3.2 Organigramme K-means clustering [82]	51
Figure 3.3 : L' algorithme de maximisation des attentes [91].....	55
Figure 3.4 : k-means [91]	55
Figure 3.5 : algorithme hiérarchique et ses variantes [91].	56
Figure 3.6 : DBSCAN et OPTICS [91].....	57
Figure 4.1 : Matrice de confusion	61
Figure 4.2 : Organigramme de notre Système CADx (Computer Aided Diagnosis).....	62
Figure 4.3 : résultats de la segmentation d'images par k-means, (a) image originale 236.png du MIAS, (b) image segmentée par k-means, (c) extraction du tumeur(masse mammaire) ...	63
Figure 4.4 : segmentation avec k valeurs ; (a), (f) originales (81B.png et 315B.png) ; (b),(c),(d) segmentation k=10,11,12 respectivement ; (f),(g),(h) segmentation k=6,7,8 respectivement ; (e),(j) résultats k-means	64
Figure 4.5 : Première fenêtre.....	65
Figure 4.6 : fenêtre d'accueil.....	66
Figure 4.7 : choix de l'image	66
Figure 4.8 : image choisie	67
Figure 4.9 : Segmentation 315B.png.....	67
Figure 4.10 : tumeur extraite	68
Figure 4.12 : représente la création de la base	70
Figure 4.13 : base CAD-LBP	70
Figure 4.14 : le chargement de l'image qui contient la région d'intérêt	70
Figure 4.15 : Application de LBP sur la région d'intérêt.....	71
Figure 4.16 : l'ajout d'un vecteur de caractéristiques dans la base.....	71
Figure 4.17 : Base d'apprentissage complète avec tous les vecteurs de caractéristiques	72

Figure 4.18 : Interface principale de WEKA	72
Figure 4.19 : Chargement de la base d'apprentissage	73
Figure 4.20 : Résultats de classification obtenus par le classifieur Bayes Network	73
Figure 4.21 : Résultats de classification obtenus par le classifieur Naive Bayes.....	74
Figure 4.22 : Résultats de classification obtenus par le classifieur Logistic	74
Figure 4.23 : Résultats de classification obtenus par le classifieur Multi Layer Perceptron ...	75
Figure 4.24 : Résultats de classification obtenus par le classifieur Simple Logistic	75
Figure 4.25 : Résultats de classification obtenus par le classifieur LWL	76
Figure 4.26 : Résultats de classification obtenus par le classifieur Ada Boost M1	76
Figure 4.27 : Résultats de classification obtenus par le classifieur Multi Class Classifieur	77
Figure 4.28 : Résultats de classification obtenus par le classifieur Multi Class Classifieur Updateable.....	77
Figure 4.29 : Résultats de classification obtenus par le classifieur Decision Table.....	78
Figure 4.30 : Résultats de classification obtenus par le classifieur Decision Stump	78
Figure 4.31 : Résultats de classification obtenus par le classifieur J48	79
Figure 4.32 : Résultats de classification obtenus par le classifieur Random Forest	79
Figure 4.33 : Résultats de classification obtenus par le classifieur Random Tree	80

Liste des Tableaux

Tableau 2.1 : Résumé des travaux proposés.....	46
Tableau 4.1 : Récapitulatif des performances des différents classifieurs.....	81

Liste des Abréviations

ANCE : Adaptive Neighborhood contraste Enhancement

ANN : Artificial Neural Networks

CAD : Computer Aided Diagnosis

CBIS : Curated Breast Imaging Subset

CCIS : Carcinome Canalaire In Situ

CIE : Codage de la couleur selon la commission Internationale de l'Eclairage

CMJ : Cyan, Magenta, Jaune

CMJN : Cyan, Magenta, Jaune, Noir pur

CNN : Convolutional Neural Network

CT : Computerized Tomography

DAO : Detection/Diagnostic Assisté par Ordinateur

DAOe : Détection Assisté par Ordinateur

DAOx : Diagnostic Assisté par Ordinateur

DDSM : Curated Breast Imaging Subset

EHO : Elephant Herding Optimization

FCM : Fuzzy C-Means

FN : Faux Negatif

FNR : False Negative Rate

FP : Faux Positif

FPR : False Positive Rate

GLCM : Gray-Level Co-occurrence Matrix

IBM : International Business Machines

IDE : Integrated Development Environment

IRM : Imaging Resonance Magnetic

LBP : Local Binary Patterns

LSTM : Long-Term Memory Network

MCC : Matthews correlation coefficient

MIAS : Mammographic Image Analysis Society

OMS : Organisation Mondiale de la Santé
PEIPA : Pilot European Image Processing Archive
PET : Position Emission Tomography
RBF : Radial Basis Functions
RL : Régression logistique
RVB : Rouge, Vert, Bleu
SGLD : Spatial Gray Level Dependency
SOM : Self Organizing Map
SVM : Support Vector Machines
TSL : Teinte, Saturation, Luminance
VN : Vrai Negatif
VP : Vrai Positif
WEKA : Waikato Environment for Knowledge Analysis

Introduction Générale :

1.Contexte et problématique :

Le cancer du sein est la maladie la plus fréquente chez les femmes dans le monde entier. Avec une détection au plutôt, Il est curable, donc cela permet d'augmenter les chances de survie des patientes.

Les lésions mammaires sont classées en calcifications ou en masses. La détection à temps des masses cancéreuses est essentiellement indispensable pour un traitement adéquat en raison de leur taille modeste lorsque les patientes ne présentent aucun symptôme initial. Les masses mammaires varient en intensité, en distribution, et en forme dans la région du sein, ce qui hausse la probabilité d'un mauvais diagnostic. Il est diagnostiqué comme maligne lorsque les tumeurs ont une forme irrégulière, des bords ambigus et des limites floues; par contre, les masses bénignes sont souvent denses, bien définies, circonscrites et grossièrement sphériques.

Vue l'hétérogénéité, la diversité morphologique, les limites floues et la taille variable des cellules cancéreuses(tumeurs malignes); il y'a une complication pour les médecins de les reconnaître, ce qui entraîne des biopsies inutiles.

La mammographie est généralement sollicitée pour l'examen de dépistage du cancer du sein. Elle permet une bonne analyse des cancers de petite taille, à un stade précoce, avant l'apparition de symptômes.

La lecture des images mammographiques est une tâche très exigeante pour les radiologues, car elle est un examen répétitif qui demande beaucoup d'attention aux détails. Elle est une tâche compliquée due à la différence d'interprétation du même examen de temps en temps, selon l'expert. Raison pour laquelle, ces dernières années, le développement de méthodes de traitement des images mammographiques est devenu intéressant, pour augmenter la précision du diagnostic et apporter aux radiologues des conseils supplémentaires et les aider à minimiser certaines erreurs. Ces méthodes ont été utilisées pour développer des systèmes de détection assistée par ordinateur et des systèmes de diagnostic assisté par ordinateur .

Les systèmes de détection/diagnostic assistés par ordinateur sont très utiles pour détecter et diagnostiquer les anomalies des seins, ils peuvent servir de seconde vue dans la détection

précoce du cancer du sein, car ils aident les médecins à lire et à interpréter l'image mammographique.

Un système de diagnostic assisté par ordinateur comprend les étapes suivantes : un prétraitement, extraction de la région d'intérêt , extraction de caractéristiques et classification . Mais dans notre cas le système utilisera des images mammographiques déjà prétraitées ensuite il fera la segmentation par une méthode de segmentation courante k-means, cette image segmentée sera traitée par la méthode des motifs locaux binaires, il récupère cela et fait une extraction des caractéristiques des statistiques du premier ordre et enfin la classification ceux-ci sont les étapes de notre système.

2. Objectifs :

L'objectif de ce sujet est de concevoir un système d'aide au diagnostic médical robuste capable de différencier entre les masses bénignes et malignes dans les images mammographiques et de fournir une vue objective aux radiologues dans l'étape de diagnostic. Nous allons cerner les régions d'intérêts par une méthode de segmentation, ensuite, la méthode des motifs binaires locaux sera appliquée, et afin de ressortir l'information de texture, des caractéristiques statistiques seront utilisées. Pour l'étape de classification, nous allons choisir quelques techniques de classification connues.

3. Plan de travail :

Notre projet porte sur 4 chapitres qui sont les suivants :

Chapitre1 : parle du contexte médical en général (définition du cancer du sein, sa particularité, les causes, les conséquences, les différents remèdes ou solutions pour le soigner et comment vivre avant et après le cancer).

Chapitre2 : il donne des détails sur les systèmes d'aide au diagnostic médical CAD (histoires, description, leurs composition)

Chapitre 3 : évoque la segmentation des masses mammaires par l'algorithme k-means (histoires du k-means, son fonctionnement et son utilisation dans la segmentation et en particulier dans la segmentation des masses)

Chapitre 4 : détaille la conception et l'implémentation de notre système avec des captures d'écran du système pour l'appui.

Chapitre 1 : Contexte médical

1.1 Introduction :

Dans ce chapitre, nous parlerons du cancer de sein en général, par la suite nous évoquerons le diagnostic du cancer sein, ensuite viendra les différents traitements du cancer du sein, la surveillance après les traitements, ainsi que le comportement d'avant et après le cancer.

Vue sa complexité et sa négativité sur la vie humaine, ce chapitre nous aidera à mieux comprendre en détails ce phénomène qui détruit plusieurs vies sans être diagnostiqué d'avance.

1.2 Le cancer du sein :

Généralités :

Le cancer du sein est la deuxième cause de décès par cancer chez les femmes dans le monde [1]. Selon l'enquête du Centre international de recherche sur le cancer et l'Organisation mondiale de la santé, en 2012, le cancer du sein a tué 8,2 millions de personnes avec une moyenne de 450 000 décès dans le monde chaque année [9]. Environ 10 à 20,8 % des cas sont dus à un cancer du sein triple négatif [3]. A l'âge groupe de 30 à 60 ans, une femme sur huit aux États-Unis et quatre sur 1000 en Inde [7] reçoivent un diagnostic de cancer du sein.

En Jordanie, 37,4% par an [4], et au Maroc, 30 à 40 000 femmes ont reçu un diagnostic de cancer du sein [5]. Au cours des années 1990-2015, parmi les sept régions définies par le Health Metrics and Evaluation, la mortalité a augmenté au maximum le taux de cancer du sein, cela a été observé en Amérique latine et dans les Caraïbes [6]. Actuellement, le taux de mortalité causé par le cancer du sein a augmenté à tous les âges [5] car diagnostiqué très tardivement [8]. Ainsi, une détection précoce et précise des tumeurs malignes est très cruciale [2].

Le cancer du sein est l'un des types de cancer les plus importants qui menace la vie et entraîne la mort chez les femmes aujourd'hui [11]. Selon l'Organisation mondiale de la santé (OMS), 2,3 millions de femmes ont reçu un diagnostic de cancer du sein en 2020 et aussi 685 000 personnes sont mortes du cancer du sein la même année [10]. Par conséquent, le traitement du cancer du sein est important en termes de protection de la santé des femmes et de réduction dépenses de santé [11].

Au cours des dernières décennies, le cancer du sein est la cause la plus fréquente du décès chez les femmes dans le monde. Les rapports scientifiques montrent que le traitement précoce du cancer du sein augmente le taux de la survie de patients [12, 13]. Plusieurs modalités d'imagerie sont appliquées pour la détection du cancer du sein comme les rayons X, les ultrasons, imagerie par résonance magnétique (IRM), histologie, émission de positons de la tomographie (PET) et la tomographie informatisée (CT) [14].

1.2.1 Description du sein :

Les seins jouent un rôle important dans la féminité et dans l'image que la femme a de son corps. La fonction biologique du sein est de produire du lait afin de nourrir un nouveau-né. La structure du sein est complexe. Chaque sein (appelé aussi glande mammaire) est composé de quinze à vingt compartiments séparés par du tissu graisseux qui donne au sein la forme qu'on lui connaît. Chacun de ces compartiments est constitué de lobules et de canaux. Le rôle des lobules est de produire le lait en période d'allaitement ; les canaux transportent ensuite le lait vers le mamelon. Pour mieux visualiser cette structure, on peut imaginer un arbre avec plusieurs branches (les canaux) rattachées à un point central (le mamelon). Aux minuscules extrémités des branches se trouvent les lobules [15].

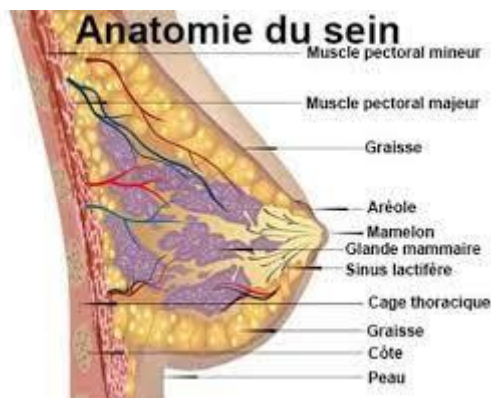


Figure 1.1 Anatomie du sein [17]

À la puberté, la jeune fille observe des changements dans la forme et le volume de ses seins sous l'influence des hormones sexuelles. Ces hormones, fabriquées par les ovaires, ont une influence sur les seins tout au long de la vie. Elles sont de deux types :

les œstrogènes permettent le développement des seins au moment de la puberté. Ces hormones sont fabriquées au cours de la première partie du cycle menstruel, après les règles ;

la progestérone est principalement présente lors de la seconde partie du cycle menstruel, avant les règles. Elle a une action complémentaire à celle des œstrogènes [15].

1.2.2 Concept cancer :

Un cancer est une maladie de la cellule. La cellule est l'unité de base de la vie. Il en existe dans le corps plus de deux cents types différents. Toutes ont un rôle précis : cellules musculaires, nerveuses, osseuses, etc. Une cellule cancéreuse est une cellule qui s'est modifiée. Habituellement, ces modifications sont réparées par l'organisme. Cependant, lorsque la cellule devient cancéreuse, elle perd ses capacités de réparation. Elle se met alors à se multiplier et finit par former une masse qu'on appelle tumeur maligne [15].

1.2.3 Définition du cancer du sein :

Un cancer du sein est une tumeur maligne qui se développe au niveau du sein. Une femme sur onze développe un cancer du sein au cours de sa vie. La vie de famille, de couple et professionnelle s'en trouve profondément modifiée [15]. C'est un cancer qui se forme dans les tissus du sein, généralement dans les conduits (tubes qui transportent le lait vers le mamelon) ou des lobules (glandes qui fabriquent le lait). Il survient chez les hommes et les femmes, bien que le cancer du sein chez l'homme soit rare [16].

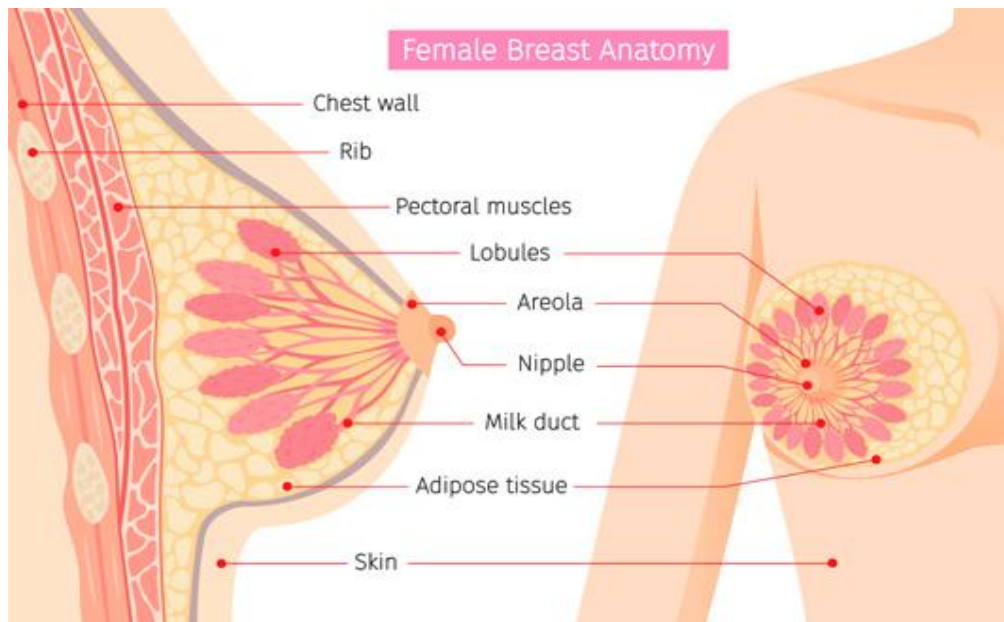


Figure 1.2 Anatomie du sein féminin [18]

Les cellules des tumeurs malignes ont tendance à quitter leur tissu d'origine et à envahir les tissus voisins; on parle de cancer infiltrant. Certaines tumeurs restent dans leur tissu

d'origine sans infiltrer les tissus voisins. On parle alors d'un cancer in situ (« resté en place »). Les cellules des cancers infiltrants peuvent se détacher de la tumeur d'origine et se propager par les vaisseaux sanguins ou lymphatiques.

Elles s'accumulent alors dans les ganglions lymphatiques voisins. On dit qu'il existe un envahissement ganglionnaire. Les cellules cancéreuses ont tendance à migrer dans d'autres organes ou parties du corps, et à y développer de nouvelles tumeurs qu'on appelle métastases. On dit dans ce cas que le cancer est métastatique.

Tous les cancers ne se comportent pas de la même façon. C'est la raison pour laquelle il est nécessaire d'envisager un traitement adapté à chaque cancer. Tous les traitements ont pour but de supprimer les cellules cancéreuses. En l'absence de traitement adapté, le cancer risque d'évoluer plus vite.

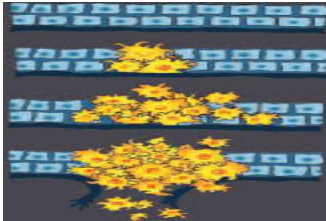


Figure 1.3 Cellules cancéreuses (cellules normales, cellules cancéreuses, cancer in situ, cancer infiltrant (Rupture de la membrane, propagation par les capillaires et vaisseaux lymphatiques)) [19]

Grâce aux résultats de la recherche, on connaît mieux aujourd'hui le cancer du sein. Ces études ont permis d'améliorer son diagnostic et son traitement.

Toutes les femmes atteintes d'un cancer du sein n'ont pas le même cancer; elles n'ont donc pas le même traitement. Le médecin propose à chaque patiente un traitement adapté à sa situation. C'est ce qui permet d'obtenir les meilleures chances de guérison [15].

1.2.3.1 Cancer fréquent :

Le cancer du sein est le cancer le plus fréquent chez la femme dans les pays occidentaux. Un quart des cancers de la femme sont des cancers du sein. En France, environ 42000 nouveaux cancers du sein sont diagnostiqués chaque année.

Le cancer du sein atteint des femmes à des âges différents, mais il se développe le plus souvent autour de 60 ans. Près de la moitié des cancers du sein sont diagnostiqués entre 50 et 69 ans. Ce cancer peut aussi apparaître chez l'homme, mais c'est extrêmement rare (moins de 1 % des cancers du sein) [15].

1.2.3.2 Les différents types de cancer du sein :

Il existe différents types de cancer du sein. Les plus fréquents (95 %) se développent à partir des cellules des canaux (cancer canalaire) et des lobules (cancer lobulaire). On les appelle des adénocarcinomes. Le cancer du sein peut être classé selon qu'il est non invasif ou invasif :

❖ *Cancer du sein non invasif (in situ)*

Le carcinome canalaire in situ (CCIS) est une lésion pré-maligne, ce n'est pas encore un cancer, mais peut évoluer pour devenir une forme invasive de cancer du sein. Dans ce type de cancer, les cellules cancéreuses se trouvent dans les conduits du sein mais n'ont pas pu se répandre dans le tissu mammaire sain.

La néoplasie lobulaire (anciennement appelée carcinome lobulaire in situ) se produit lorsqu'il y a des changements dans la paroi cellulaire les lobules, qui indiquent qu'il existe un risque accru de développer un cancer du sein à l'avenir. Le néoplasie lobulaire n'est pas réellement un cancer du sein, et bien que les femmes atteintes de néoplasie lobulaire aient des examens réguliers, la plupart ne développera pas de cancer du sein.

❖ *Cancer du sein invasif :*

Le cancer du sein invasif est le nom donné à un cancer qui s'est propagé à l'extérieur des canaux (cancer canalaire invasif cancer) ou des lobules (cancer du sein lobulaire invasif). Ceux-ci peuvent être encore classifiés par leur histologie; pour par exemple, les tumeurs mammaires tubulaires, mucineuses, médullaires et papillaires sont des sous-types plus rares de cancer du sein.

On distingue plusieurs situations :

Lorsque les cellules cancéreuses ont infiltré le tissu qui entoure les canaux et les lobules, on parle de cancer ou carcinome infiltrant. Le cancer canalaire est le cancer infiltrant le plus fréquent. Huit cancers sur dix sont des cancers canaux infiltrants. Le cancer lobulaire infiltrant est plus rare. Les cancers infiltrants peuvent se propager vers les ganglions ou vers d'autres parties du corps. Les ganglions les plus souvent atteints par les cellules du cancer du sein se trouvent sous les bras, au niveau des aisselles ; on les appelle les ganglions axillaires. Le traitement des cancers infiltrants repose sur un traitement local (chirurgie, radiothérapie), associé dans certains cas à un traitement général (chimiothérapie, hormonothérapie). Lorsque les cellules cancéreuses se trouvent uniquement à l'intérieur des canaux ou des lobules, sans que la tumeur ait infiltré le tissu qui les entoure, on parle de cancer in situ :

le cancer canalaire in situ ou carcinome canalaire in situ est le plus fréquent : huit cancers in situ sur dix sont des cancers canaux in situ. Ils sont le plus souvent découverts lors d'un dépistage systématique (une mammographie ou un examen des seins par le médecin) et

quelquefois révélés par un écoulement de sang du mamelon ou un eczéma du mamelon. Il arrive que les cancers canaux in situ infiltrent les tissus voisins et deviennent des cancers infiltrants. Le traitement des cancers canaux in situ est un traitement local comme la chirurgie ou la radiothérapie

le cancer lobulaire in situ (ou néoplasie lobulaire in situ) est plus rare. Il est plutôt considéré comme un facteur de risque de développer un cancer du sein que comme un cancer. Sa prise en charge est différente des autres cancers du sein.

Chaque cancer est particulier et nécessite un traitement approprié [15].

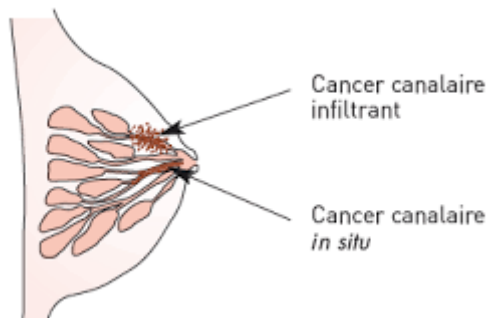


Figure 1.4 Cancer canalaire infiltrant et cancer canalaire in situ [20]

1.2.4 Quelques symptômes du cancer du sein :

Les symptômes du cancer du sein comprennent :

- Une grosseur dans le sein
- Modification de la taille ou de la forme du sein
- Capitons de la peau ou épaissement du tissu mammaire
- Un mamelon inversé
- Éruption cutanée sur le mamelon
- Écoulement du mamelon
- Gonflement ou bosse sous l'aisselle
- Douleur ou inconfort dans le sein qui ne disparaît pas
- Rougeur de la peau
- Épaississement de la peau [15]

1.2.5 Les causes d'un cancer du sein :

Certaines femmes atteintes d'un cancer du sein cherchent à un moment donné de leur maladie à connaître les causes de leur cancer. Les scientifiques cherchent également à connaître les raisons de son apparition. Malgré les progrès qui ont permis de mieux connaître les mécanismes de développement des cancers, les causes du cancer du sein ne sont actuellement

pas connues. Néanmoins, les études ont mis en évidence certains facteurs de risque qui favorisent un cancer du sein.

On distingue différents types de facteurs de risque.

L'âge est un facteur important. Comme pour la plupart des maladies cardiovasculaires et la plupart des cancers, le risque d'avoir un cancer du sein augmente avec l'âge : le vieillissement favorise l'apparition des cancers. Pour ces raisons, des programmes de dépistage du cancer du sein ont été mis en place pour les femmes à partir de 50 ans.

Certains facteurs de risque ont été identifiés au plan collectif : des premières règles avant 12 ans et une ménopause après 55 ans ; une absence de grossesse ou une première grossesse après 40 ans ; un traitement hormonal substitutif de la ménopause prescrit pendant plus de 10 ans ; une consommation exagérée d'alcool, de sucres et de graisses d'origine animale ainsi que l'obésité.

En revanche, le nombre de grossesses et l'allaitement prolongé pourraient diminuer le risque de développer un cancer du sein. Les études scientifiques n'ont pas trouvé de lien entre : l'utilisation de déodorants ou anti-transpirants à base de sels d'aluminium et le développement d'un cancer du sein; un choc sur un sein et le développement d'un cancer. Cependant, une femme peut découvrir une tumeur du sein après s'être heurtée parce qu'elle est plus attentive à son sein. Les données d'aujourd'hui disponibles ne permettent pas de conclure à l'existence d'une relation entre l'apparition d'un cancer du sein et un stress ou un événement douloureux (comme une séparation, un deuil, la perte d'un d'emploi, etc.). Des études continuent pour mieux connaître ces facteurs et en identifier d'autres.

Les facteurs familiaux ou génétiques : Un petit nombre de cancers du sein est lié à des facteurs familiaux ou génétiques. C'est parfois le cas lorsque : plusieurs femmes de la même famille ont eu ou ont un cancer du sein ; une femme de la même famille a eu un cancer du sein avant l'âge de 40 ans ; plusieurs personnes de la même famille ont eu ou ont un cancer du côlon ou des ovaires.

Dans ce cas, le médecin peut proposer une consultation chez un spécialiste d'oncologie génétique. Même si la présence d'un ou de plusieurs de ces facteurs favorise le développement d'un cancer du sein, il faut savoir que, mis à part les facteurs familiaux ou génétiques, ces facteurs augmentent généralement assez peu le risque : une femme qui possède une ou plusieurs de ces caractéristiques peut ne jamais développer un cancer du sein. En revanche, il est possible qu'une femme n'ayant aucune de ces caractéristiques soit atteinte d'un cancer du sein. Les facteurs de risque ne permettent donc pas de déterminer les causes exactes à l'origine d'un cancer du sein chez une femme [15].

La cause précise du cancer du sein est inconnue, mais plusieurs facteurs de risque de développer la maladie ont été identifiés. Il est important de se rappeler que le fait d'avoir un facteur de risque augmente le risque de développer un cancer, mais il ne signifie pas que vous aurez définitivement un cancer. De même, ne pas avoir de facteur de risque ne signifie pas que vous n'aura pas de cancer.

Facteurs de risque les plus importants : Genre féminin; Exposition aux rayonnements ionisants; Vieillir; Avoir moins d'enfants; Predisposition génétique (antécédents familiaux ou mutations de certains gènes); Antécédents d'hyperplasie atypique; Obésité ; Exposition aux œstrogènes; Alcool. Il existe divers facteurs de risque associés au développement du cancer du sein, bien que la plupart des facteurs ne s'applique pas à toutes les femmes qui développent la maladie [15].

1.3 Le diagnostic du cancer du sein :

Différentes étapes sont nécessaires pour établir ou non le diagnostic de cancer du sein.

1.3.1 Comment découvre-t-on un cancer du sein :

Le cancer du sein est généralement diagnostiqué par examen clinique, imagerie et biopsie. Un cancer du sein est le plus souvent diagnostiqué à quatre occasions :

Lors de la découverte de symptômes par la patiente elle-même. Lorsqu'une femme découvre une boule ou une anomalie au niveau du sein (écoulement par le mamelon, mamelon rétracté, asymétrie ou ride) ou dans une aisselle, il est important de consulter un médecin pour qu'il réalise les examens nécessaires. Les études scientifiques montrent qu'un cancer découvert tôt a de meilleures chances de guérison. Le délai entre la découverte de quelque chose d'anormal et le début d'un traitement peut parfois sembler très long. Ce temps est nécessaire pour réaliser les examens, établir le diagnostic et définir le traitement le mieux adapté à chaque situation.

Lors d'une consultation de dépistage, lors d'une consultation ou d'un bilan de santé, le médecin généraliste ou le gynécologue proposent de chercher une anomalie du sein après avoir expliqué les raisons et les conséquences éventuelles de ce dépistage. Lors d'une consultation habituelle chez le gynécologue, lors d'une visite de contrôle, le médecin peut trouver une anomalie au niveau des seins. Lors de la surveillance d'un premier cancer du sein, Lors de la surveillance d'un cancer du sein traité, le médecin vérifie qu'un second cancer du sein ne s'est pas développé [15].

1.3.2 Le dépistage :

Un dépistage consiste à détecter un cancer avant qu'il ne soit palpable ou qu'il ne se traduise par un signe anormal comme une modification de la peau ou du mamelon.

Détecter tôt certains cancers, permet de mieux les traiter en proposant des traitements moins lourds et qui offrent plus de chances de guérison. Le but du dépistage est de détecter parmi des personnes à priori non malades celles qui présentent des anomalies susceptibles d'être cancéreuses ou d'évoluer en cancer afin de les traiter rapidement [15]. L'examen utilisé pour dépister un cancer du sein est une mammographie (radiographie des seins).

Une mammographie détecte des anomalies de petite taille, dont certaines seulement se révéleront être un cancer. Ces anomalies sont parfois détectées même si l'examen clinique est normal.

Si une anomalie est découverte, le médecin prescrit des examens complémentaires (mammographie complémentaire, échographie, ponction et éventuellement biopsie) afin de confirmer ou d'éliminer le diagnostic de cancer [15].

Près de 75 % des cancers du sein sont diagnostiqués entre 50 ans et 69 ans. C'est la raison pour laquelle l'État et la Caisse nationale d'Assurance Maladie ont mis en place un programme national de dépistage des cancers du sein. On parle aussi de dépistage organisé du cancer du sein. Ce programme offre à toutes les femmes de 50 à 74 ans une mammographie de dépistage gratuite tous les deux ans. La majorité des femmes se fait surveiller mais trop de femmes ne le font pas régulièrement [15].

À partir de 50 ans, un examen médical des seins tous les ans ainsi qu'une mammographie tous les deux ans sont conseillés. Dans le cadre du programme national de dépistage des cancers du sein, la mammographie de dépistage est réalisée dans des conditions spécifiques : elle est pratiquée par des radiologues qui doivent avoir reçu une formation spéciale; pour plus de sécurité, les images obtenues par la mammographie sont interprétées une deuxième fois par un second radiologue également spécialisé en mammographie [15].

1.3.3 Fiabilité du diagnostic de cancer du sein :

Le bilan diagnostique confirme que l'anomalie suspecte est bien un cancer. Il précise le type de cancer, évalue sa taille et détermine s'il s'est développé dans d'autres parties du corps. Le diagnostic peut être réalisé par différents médecins spécialistes : gynécologue, sénologue, cancérologue ou encore médecin radiologue lors d'une mammographie.

Dans un premier temps, le médecin interroge la patiente sur ses antécédents médicaux, chirurgicaux et gynéco-obstétricaux, personnels et familiaux. Il lui pose des questions sur ses symptômes actuels qui le guident dans son diagnostic. Il est important de mentionner tous les symptômes repérés même s'ils ne correspondent pas systématiquement à un cancer. Après l'entretien, le médecin réalise un examen physique approfondi et complet des seins. L'examen physique prend en compte huit éléments :

- ❖ **La taille de la tumeur :** Le médecin demande à la patiente de s'allonger sur le dos. Le bras est relevé du côté du sein malade. Dans cette position, les seins sont plus étalés et plus faciles à palper. Si l'anomalie est palpable, le médecin évalue sa taille.
- ❖ **La mobilité de la tumeur :** Le médecin regarde ensuite si la tumeur bouge sous la peau ou si elle est fixée à la paroi thoracique ou à la peau.
- ❖ **La localisation de la tumeur :** Il est important de repérer l'endroit où la tumeur est localisée dans le sein. Cette information est nécessaire pour le choix des traitements. On distingue quatre secteurs différents [15].

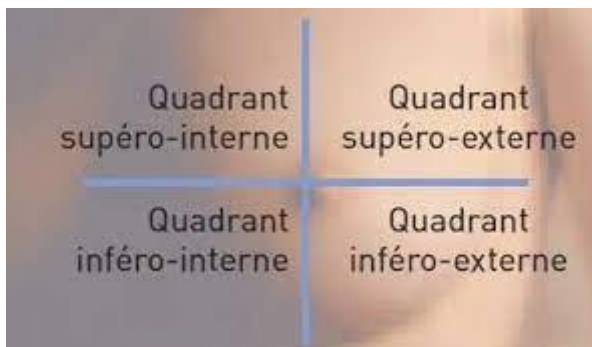


Figure 1.5 Les quatre secteurs du sein [21]

- ❖ **L'aspect de la peau :** On observe parfois une modification de la peau autour de la tumeur. Cette modification est variable : la peau peut être rouge (c'est ce qu'on appelle un érythème), écorchée (ulcérée), gonflée (on dit alors qu'il y a un œdème) ou encore être recouverte de petites boules dures (nodules). Elle peut également être plissée ou ridée.
- ❖ **La forme du mamelon et de l'aréole :** Le médecin recherche systématiquement s'il existe une déformation, une rétraction, une ulcération du mamelon ou de l'aréole. Ces modifications sont parfois accompagnées d'un écoulement au niveau du mamelon.
- ❖ **L'augmentation de la taille de la tumeur :** Le médecin demande à la patiente si elle a remarqué une augmentation de la taille de la tumeur.

- ❖ **La palpation des ganglions :** Le cancer du sein s'accompagne aussi parfois d'une augmentation de la taille des ganglions au voisinage du sein (dans ce cas, on parle d'adénopathie). Le médecin recherche systématiquement les ganglions anormaux en palpant les différents endroits où ils peuvent se trouver (essentiellement dans l'aisselle).
- ❖ **Les examens complémentaires.** En fonction de l'entretien et de l'examen physique, le médecin demande parfois des examens complémentaires pour confirmer le diagnostic [15].

1.3.4 Les examens complémentaires :

Les examens complémentaires les plus fréquents du bilan diagnostique du cancer du sein sont : les examens radiologiques (mammographie des deux seins, échographie mammaire) ; les prélèvements (cytologie, histologie) ; les examens sanguins.

Certains examens sont systématiques ; d'autres dépendent de l'anomalie découverte et de l'état général de la personne. Les examens complémentaires ont trois fonctions essentielles : confirmer le diagnostic ; déterminer si des cellules cancéreuses se sont propagées dans d'autres parties du corps ; aider à choisir le traitement le mieux adapté.

En fonction des résultats, le médecin prescrit d'autres examens : radiographie du thorax, scintigraphie osseuse, scanner, échographie abdominale ou IRM. Ces examens ainsi que les examens sanguins ne sont pas réalisés de façon systématique, mais selon les résultats de l'examen clinique, de la mammographie et de l'histologie [15].

1.3.4.1 Les examens radiologiques :

Les examens radiologiques permettent d'obtenir des images d'une partie du corps ou des organes à l'aide de différents appareils. On les appelle aussi examens d'imagerie.

La mammographie est un examen radiologique qui montre des images de l'intérieur du sein. Cet examen est systématique lors du bilan, il aide le médecin à établir le diagnostic. C'est la raison pour laquelle la qualité de la mammographie est essentielle. Cet examen permet de distinguer différents types d'anomalies qui ne sont pas toujours des cancers.

La mammographie est complétée par une échographie mammaire si besoin, notamment pour recueillir des images qui précisent celles obtenues préalablement par la mammographie ou pour guider un prélèvement [15].

1.3.4.2 Les prélèvements :

Tous les cancers ne sont pas détectés par une mammographie ou une échographie mammaire. En présence d'une boule dans le sein ou d'une anomalie décelée sur les images de la mammographie ou de l'échographie, on a recours à un prélèvement afin de réaliser un examen au microscope : un examen anatomopathologique. L'examen anatomopathologique n'est pas un traitement. Il donne des informations précises sur le type de cancer du sein et ses caractéristiques grâce à l'analyse au microscope des cellules et des tissus prélevés au niveau d'une anomalie du sein.

On distingue deux types d'examens au microscope : la cytologie (examen cytopathologique) et l'histologie (examen histopathologique). Ils sont réalisés par un médecin spécialiste, l'anatomopathologiste. Seul l'examen histopathologique élimine ou affirme de façon définitive le diagnostic de cancer [15].

1.3.4.3 Les examens sanguins :

Lors de la première consultation, le médecin ne prescrit pas systématiquement des prises de sang. Elles sont réalisées selon les résultats des autres examens (examen clinique, examens radiologiques, examen anatomopathologique) et en fonction des traitements envisagés.

Elles sont alors effectuées avant de débiter tout traitement. Les cellules du corps libèrent parfois dans le sang des substances particulières. Les cellules de certains cancers libèrent des quantités plus ou moins importantes de ces substances, appelées alors marqueurs tumoraux. Le dosage des marqueurs tumoraux se fait grâce à une prise de sang. Il peut donner des indications sur l'évolution de la maladie.

Les études scientifiques montrent que le dosage des marqueurs tumoraux n'a pas d'utilité dans le diagnostic des cancers du sein non métastatiques. L'ensemble des examens du bilan diagnostique permet d'élaborer un plan thérapeutique, c'est-à-dire de proposer à la patiente le ou les traitements adaptés à sa situation et la meilleure façon de les réaliser. Il existe plusieurs formes de cancer du sein : toutes les femmes atteintes d'un cancer du sein n'ont pas le même traitement. Car elles n'ont pas le même cancer. À chaque situation particulière correspond un traitement particulier adapté. C'est ce qui permet d'obtenir les meilleures chances de guérison [15].

1.4 Les différents traitements du cancer du sein :

Les études scientifiques permettent d'évaluer les nouveaux traitements ou les nouvelles associations de traitements, et de voir dans quel ordre ils sont le plus efficaces. Ces études permettent également de comparer leurs avantages et leurs inconvénients par rapport aux traitements habituels.

Lorsque ces études ont montré que, pour une situation donnée, il existe un traitement qui présente plus d'avantages par rapport aux autres, on parle de traitement standard. Il s'agit du traitement de référence. Le traitement standard est alors proposé de façon systématique dans cette situation. Il arrive cependant que le médecin ne puisse pas l'appliquer du fait de facteurs particuliers liés à la patiente ou à sa maladie. Le médecin propose alors un ou plusieurs traitements mieux adaptés à la situation.

Souvent, plusieurs traitements ont des bénéfices et des inconvénients comparables. Lorsque les études scientifiques n'ont pas pu identifier un traitement dont les avantages et les inconvénients sont préférables à ceux des autres traitements, plusieurs possibilités existent alors : ce sont des options [15].

Les traitements du cancer du sein consistent à enlever la tumeur et à supprimer toutes les cellules cancéreuses. Ils ont pour but de : guérir la personne ; réduire le risque de récurrence ; augmenter la durée de vie de la personne malade ; améliorer la qualité de vie. Différents types de traitements sont effectués seuls ou associés entre eux :

➤ La chirurgie :

Consiste lors d'une opération à enlever la tumeur et les éventuelles greffes cancéreuses, c'est-à-dire les endroits que les cellules cancéreuses auraient pu atteindre. La chirurgie est le traitement le plus anciennement utilisé pour soigner les cancers du sein. C'est un traitement standard. La chirurgie du cancer du sein a quatre objectifs : confirmer le diagnostic et préciser le stade d'évolution du cancer, notamment examiner si les ganglions ont été atteints par des cellules cancéreuses ; enlever la tumeur ; prélever et examiner certains ganglions ; conserver ou restaurer la taille et la forme du sein après l'ablation de la tumeur ou de la totalité du sein. Deux types de chirurgie peuvent être pratiqués : une chirurgie conservatrice ou une mastectomie (mammectomie) totale [15].

La chirurgie conservatrice : a pour but d'enlever la totalité de la tumeur sans enlever la totalité du sein. On parle aussi de tumorectomie, d'exérèse, ou encore de mastectomie partielle. La tumeur est enlevée avec une large marge de tissu autour d'elle. Cette marge de sécurité limite

ainsi le risque de récurrence. Le mamelon et l'aréole sont conservés (sauf pour certaines tumeurs, situées derrière le mamelon et l'aréole).

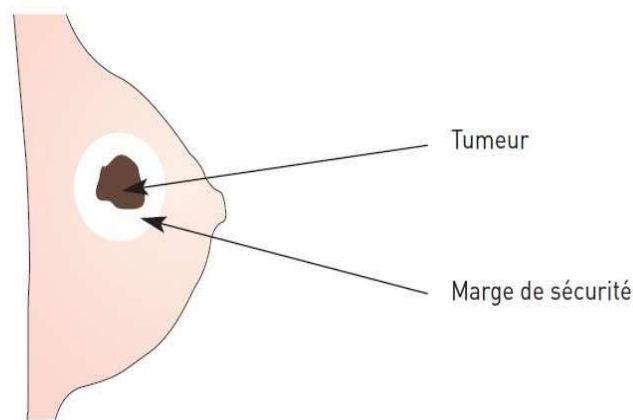


Figure 1.6 Chirurgie conservatrice [22]

Cette technique n'entraîne pas une déformation du sein trop importante. La chirurgie conservatrice s'accompagne toujours d'une radiothérapie.

La mastectomie totale : consiste à enlever la glande mammaire dans son intégralité y compris l'aréole et le mamelon. Après cette intervention, le chirurgien prescrit une prothèse mammaire externe. La reconstruction est possible environ un an après la fin du traitement de radiothérapie. Parfois, elle est réalisée en même temps que la mastectomie (reconstruction immédiate).

Les principales indications d'une mastectomie :

- Plusieurs tumeurs dans le même sein.
- Un cancer qui présente des signes d'inflammation d'abord traité par une chimiothérapie néoadjuvante.
- Un cancer volumineux par rapport au volume du sein sans possibilité de faire une chimiothérapie néoadjuvante.
- Une lésion susceptible de se transformer en cancer étendu.
- Une récurrence.
- Une contre-indication à réaliser une radiothérapie (radiothérapie du thorax effectuée dans le passé).

Lorsque la tumeur est de taille plus importante, une mastectomie totale est souvent nécessaire [15].

➤ **La radiothérapie :**

Utilise des rayons X pour détruire les cellules cancéreuses situées au niveau du sein ou dans certains ganglions. La chirurgie et la radiothérapie agissent localement sur les cellules cancéreuses situées dans le sein ou dans les ganglions : ce sont des traitements locaux du cancer. La radiothérapie est utilisée depuis de longues années pour traiter différents cancers. Elle consiste à recourir à des rayons X qui atteignent la tumeur et détruisent les cellules cancéreuses. Selon la zone et les organes de voisinage à traiter, les rayons utilisés sont différents (photons ou électrons). Ces divers types de rayons sont parfois associés entre eux. C'est pour cette raison que, pour un même traitement, une patiente peut être placée sous différents appareils : appareils de cobalthérapie ou accélérateurs linéaires. On effectue une radiothérapie en fonction du type de cancer et de son stade d'évolution. La radiothérapie est le plus souvent proposée après une chirurgie du sein (radiothérapie postopératoire). Elle a pour but de diminuer les risques que le cancer revienne localement au niveau de la paroi du thorax, de la cicatrice, du sein ou de certains ganglions, quel que soit le stade de la maladie. La radiothérapie est le plus souvent réalisée 3 à 6 semaines après la chirurgie. Un traitement par radiothérapie traite et contrôle le cancer localement en évitant le plus possible les effets secondaires et les séquelles. Il doit garantir à la patiente des résultats esthétiques satisfaisants [15].

➤ **La chimiothérapie :**

Consiste à utiliser des médicaments (par injection dans une veine le plus souvent) qui agissent sur toutes les cellules cancéreuses, même sur celles qui n'ont pas été détectées par les examens d'imagerie. La chimiothérapie est l'un des traitements du cancer du sein. Il s'agit d'un traitement dit adjuvant, c'est-à-dire qui complète un traitement chirurgical. La chimiothérapie adjuvante a pour objectif de diminuer le risque que des cellules cancéreuses se développent à distance, ce qu'on appelle des métastases.

La chimiothérapie agit sur le cancer à l'aide de médicaments appelés médicaments anti-tumoraux ou médicaments anticancéreux. Ces médicaments agissent par voie générale: ils agissent sur les cellules cancéreuses dans l'ensemble du corps, soit en les détruisant, soit en les empêchant de se multiplier.

La chimiothérapie adjuvante entraîne certains effets secondaires : nausées et vomissements, diarrhée, constipation, aphtes, sécheresse dans la bouche, chute des cheveux, effets sur le sang, fatigue, douleurs, cycles irréguliers ou arrêt des règles (passagers ou non), perturbations de la vie de couple, etc. [15].

➤ L'hormonothérapie :

Empêche l'action des hormones féminines, les œstrogènes, susceptibles de stimuler la croissance des cellules cancéreuses. La chimiothérapie et l'hormonothérapie agissent sur l'ensemble du corps : ce sont des traitements du cancer par voie générale.

La radiothérapie, la chimiothérapie et l'hormonothérapie sont souvent utilisées en complément de la chirurgie pour améliorer les chances de guérison. On les appelle alors des traitements adjuvants. Si un traitement est effectué avant une chirurgie, on parle de traitement néoadjuvant.

L'hormonothérapie est l'un des traitements du cancer du sein. C'est un traitement général qui agit dans l'ensemble du corps.

Certaines hormones secrétées par les ovaires stimulent la croissance des cellules cancéreuses. Une hormonothérapie vise à empêcher l'action de ces hormones ou à diminuer leur sécrétion afin de ralentir ou de stopper la croissance des cellules cancéreuses. Il s'agit d'un traitement adjuvant qui complète le traitement local. L'hormonothérapie du cancer du sein a deux objectifs :

- réduire le risque de récurrence dans le sein opéré, dans l'autre sein ou à distance (métastases) ;
- faciliter le traitement local du cancer. Lorsqu'une hormonothérapie est prescrite avant une chirurgie ou une radiothérapie, elle permet de diminuer la taille de la tumeur et facilite ainsi la chirurgie ou la radiothérapie. On parle alors d'hormonothérapie néoadjuvante.

Tous les cancers ne réagissent pas à l'action d'une hormonothérapie. Hormonosensible, c'est-à-dire que les cellules de la tumeur doivent posséder des récepteurs hormonaux. Ces récepteurs détectent et captent les hormones qui passent dans le sang. Un examen au microscope détermine si les cellules cancéreuses possèdent des récepteurs [15].

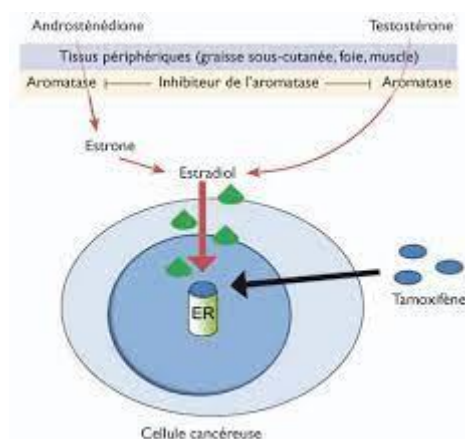


Figure 1.7 : Récepteurs hormonaux d'un cancer hormonosensible [23]

1.5 La surveillance après le traitement :

Les traitements du cancer du sein ont pour buts de soigner le cancer et de réduire le risque de récurrences locales ou à distance (métastases). Le risque de récurrence est très variable et étroitement lié au stade d'évolution du cancer au moment de son diagnostic. La plupart des récurrences du cancer du sein surviennent dans les 5 ans qui suivent le traitement. Parfois, ces récurrences sont beaucoup plus tardives. Cependant, ces récurrences ne sont pas systématiques. En effet, dans la plupart des cas, le cancer ne revient jamais. Le rôle de la surveillance est de :

- détecter des signes de récurrences de la maladie, ce qui permet de mettre en place un traitement adapté. La surveillance doit être régulière et prolongée ;
- évaluer les résultats du traitement notamment au niveau esthétique ;
- prévenir et traiter d'éventuels effets secondaires. Les effets secondaires dépendent des traitements reçus, des doses administrées et de la façon dont la femme a réagi à la maladie et aux traitements. Ils n'apparaissent pas systématiquement ;
- évaluer la qualité de vie de la patiente, notamment en fonction des traitements reçus. Chaque patiente réagit différemment en fonction de sa personnalité, de ses conditions de vie, de sa maladie, des traitements et des effets secondaires. La fatigue a un impact important sur la qualité de vie. Elle est liée au traitement ou à une réaction dépressive. La fatigue et la douleur doivent toujours être prises en compte par le médecin ;

La surveillance consiste à consulter le médecin à un rythme déterminé. Les consultations comprennent un interrogatoire, suivi d'un examen physique [15].

1.6 Vivre pendant et après un cancer :

Dans la vie d'une personne, la maladie est un moment difficile. Elle engendre des émotions particulièrement intenses chez le malade et ses proches. Chaque étape de la maladie génère des sentiments différents. Les exprimer et les partager avec d'autres permet de mieux les vivre.

Un accompagnement social doit être proposé systématiquement pour aider la patiente et ses proches. La réinsertion sociale a pour but d'aider la patiente et sa famille dans sa vie quotidienne. Tout au long de la maladie, un suivi efficace de l'assistante sociale permet à la patiente d'éviter ou de résoudre certaines difficultés pendant les hospitalisations, mais aussi après les traitements [15].

1.7 Conclusion :

Nous pouvons conclure que le cancer du sein est une maladie très fréquente dans le monde actuel chez les femmes. Une importante étude est lancée sur cette maladie en vue d'amener de nouvelles stratégies très efficaces et bénéfiques. Elle a plusieurs conséquences sur la vie d'une femme atteinte, nous pouvons constater qu'il y a différents types comme malignes, bénignes etc... avec les nouvelles technologies développées et de vastes recherches ce fléau est en voie de traitement. Des recherches sont en cours pour remédier à cette maladie et la prévenir afin de pouvoir diagnostiquer au plus tôt avant que la tumeur ne prenne de l'ampleur et la retirer au plus vite d'où l'idée de cette conception pour venir à bout de cancer.

Chapitre 2 : Les systèmes d'aide au diagnostic médical

2.1 Introduction :

Tout d'abord nous donnerons toute une petite description des systèmes d'aide au diagnostic médical. Après un détail complet sur les compositions d'un système d'aide au diagnostic assisté par ordinateur avec les différents méthodes de segmentation en vue. L'analyse de forme et la classification font partie d'une chaîne compacte indissociable de traitement numérique et automatique (ou semi-automatique) intitulée le diagnostic assisté par ordinateur (DAOx). Ainsi, une bonne évaluation de la performance d'une telle partie de description ou de classification n'exige la maîtrise de toute la chaîne de diagnostic.

Dans ce cadre, on présente dans ce chapitre un état de l'art concernant la phase de prétraitement des images mammographiques (mais dans notre cas nous utiliserons des images déjà prétraitées), la segmentation des lésions mammaires, l'extraction des caractéristiques et la classification. Cette littérature permet, d'abord, une meilleure compréhension des choix des méthodes de prétraitement et de segmentation adoptées par la suite. Enfin, l'étude des méthodes de classification permet de sélectionner les classificateurs les plus pertinents et de comparer leur performance afin de préserver le classificateur le plus approprié à notre domaine d'application [39].

2.2 Description des systèmes de Détection/Diagnostic Assistés par Ordinateur (DAO) :

2.2.1 Importance et nécessité des systèmes DAO :

Les premiers travaux concernant les systèmes automatiques de traitement des images mammographiques avaient pour but d'offrir une seconde interprétation aux radiologues afin de les aider à détecter/diagnostiquer à un stade précoce, les lésions malignes quelles que soient leurs natures masses ou microcalcifications. On les appelle les systèmes de détection/Diagnostic Assisté par Ordinateur (DAO). Les systèmes de détection Assisté par Ordinateur couramment notés (DAOe) servent à détecter et à localiser les lésions dans les images mammographiques [39]. Les systèmes de Diagnostic Assisté par Ordinateur couramment notés (DAOx) [37] désignent un système complet de traitement d'images mammographiques allant du prétraitement jusqu'à la classification et la prise de décision. Le succès de tels systèmes est dû à leur rapidité, leur consistance et leur capacité à fournir des

solutions fiables pour assister l'étape de détection des lésions mammaires ou bien l'étape d'identification [39].

2.2.2 Détection Assistée par Ordinateur (DAOe) :

Plusieurs études ont montré que les radiologues manquent entre 4% et 38% de détection de cancers [35] et que ce taux s'améliore de 15% en utilisant une deuxième lecture vue que l'interprétation est souvent difficile et dépend de l'expertise du radiologue [36]. Toutefois, il n'est pas toujours envisageable de mettre à la disposition plus d'un radiologue pour la lecture de la même mammographie à cause essentiellement des contraintes financières. Il est alors possible qu'un radiologue fasse la lecture sans assistance et par la suite, le système de l'aide à la détection (Détection Assisté par Ordinateur) lui confirme (ou non) les lésions détectées lors de l'utilisation d'algorithmes appropriés [39].

Les systèmes d'aide à la détection par ordinateur appliqués à la mammographie sont actuellement utilisés en mammographie diagnostique, mais leur place en mammographie de dépistage n'a pas encore été établie [39]. Dans le cadre d'un test national de compétence des mammographies de dépistage en Italie, [40] ont comparé le résultat de lecture d'images mammographiques par radiologues au résultat de lecture par le biais de système DAOe. Les auteurs ont conclu que la performance d'une seule lecture avec un système DAOe est similaire à celle de la lecture par deux radiologues. Krupinski et Nishikawa ont comparé les performances d'un système de détection automatique avec celles de six radiologues pour la détection des microcalcifications [41]. Ils ont constaté qu'un radiologue expérimenté ne balaye pas l'image entière mais se concentre sur les zones identifiées lors de l'évaluation initiale globale comme contenant probablement des lésions [39].

2.2.3 Diagnostic Assisté par Ordinateur (DAOx) :

La classification des lésions mammaires par un radiologue est une classification humaine subjective qui peut facilement classifier la même lésion de deux manières différentes.

En revanche, un système automatique de Diagnostic Assisté par Ordinateur (DAOx), basé sur les descripteurs (caractéristiques) de la lésion donne toujours les mêmes résultats pour le même cliché et peut donc servir de référence [39].

Un système automatique de Diagnostic Assisté par Ordinateur sert à donner un second avis au radiologue. Un tel système commence généralement par un prétraitement afin d'obtenir une image plus lisible et moins bruitée. Cette opération a pour but l'élimination de l'effet du speckle qui se présente habituellement dans les images radiologiques et échographiques. Ce

prétraitement facilite la tâche de l'étape suivante qui consiste à détecter tous types d'anomalies (image normale, présence de masse(s), présence de microcalcification (s), présence des deux types d'anomalies). Ensuite, une étape d'extraction des descripteurs mathématiques qui décrivent aux mieux les spécificités de ces lésions sont effectuées. Finalement, une étape de classification et de décision à base d'algorithmes appropriés est adoptée. Ces différentes étapes sont plus détaillées par la suite [39].

Par ailleurs, il a été démontré qu'après avoir subi une biopsie, seulement moins d'un tiers des mammographies suspectes sont prouvées êtres cancéreux. Ainsi, concevoir un système DAOx permet d'aider le radiologue à diminuer le nombre de biopsies inutiles et de recommander dans le cas contraire un suivi à court terme uniquement. En effet, il est possible d'augmenter la sensibilité de la mammographie en utilisant les systèmes DAO [39]. Ainsi, une étude détaillée concernant la cause des cancers manqués (Faux négatifs) [38] a montré que le nombre de cancers mal interprétés (19 cas) dépasse le nombre de cancers non détectés en mammographie (11 cas). Pour pouvoir diminuer le nombre de biopsies inutiles, il faut réduire le nombre des faux positifs (bénins considérés malins) ce qui est très difficile quand on veut maintenir un taux élevé de sensibilité (taux élevé d'identification des cancers). Bien que plusieurs études soient investies pour la réussite des systèmes de diagnostic assisté par ordinateur, aucune méthode n'a été commercialisée jusqu'à l'heure actuelle [39]. Notre travail sera basé sur les DAOx.

2.3 Composition des systèmes DAOx :

En pratique, le système de diagnostic assisté par ordinateur (DAOx) dédié à l'analyse d'images mammographiques est une suite de phases qui doivent être exécutées l'une après l'autre, depuis l'acquisition de l'image jusqu'à la prise de décision. Certaines de ces phases sont souvent étroitement liées et indissociables. Les étapes de traitement d'une image mammographique peuvent se résumer en :

- une étape de prétraitement qui sert à améliorer la qualité de l'image avant toutes manipulations.
- une étape de segmentation qui permet de détecter la lésion à étudier
- une étape de description qui a pour but de caractériser les lésions à travers des formulations mathématiques.
- une étape de classification et de prise de décision en utilisant un classifieur adéquat. Ces différentes étapes sont résumées dans le diagramme représenté dans la figure 2.1. On détaille dans les sections qui suivent l'état de l'art concernant ces différentes étapes d'un système DAOx [39].

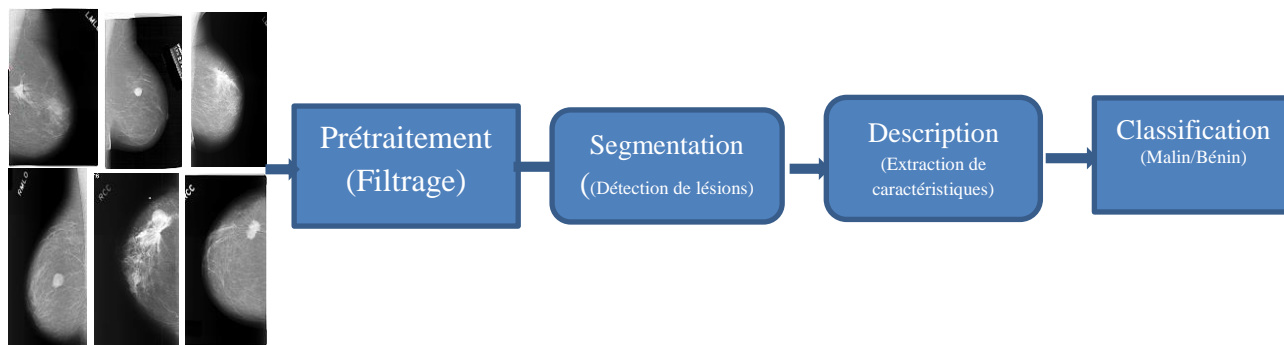


Figure 2.1 – Diagramme général d'un système DAOx [39]

2.3.1 Prétraitement :

Divers facteurs motivent les chercheurs pour développer des systèmes de diagnostic/détection par ordinateur. Le cancer du sein (comme tous les cancers d'une manière générale) doit être détecté dans sa première phase pour maximiser les chances de survie. Sauf que dans cette phase, il s'avère très difficile de repérer, à l'œil nu, la pathologie dans le tissu mammaire environnant sans avoir recours à un prétraitement spécifique de l'image acquise. D'où, l'objectif principal de cette étape est d'augmenter le contraste entre la lésion mammaire (que ce soit masse ou microcalcification) et le reste de l'image pour faciliter les traitements ultérieurs [39]. Sachant que dans le cas où une région d'intérêt diffère en luminance de moins de 2% du reste de l'image, elle demeure indiscernable à l'œil nu [42]. Le prétraitement des images mammographiques est connu sous le nom de rehaussement ou d'amélioration du contraste.

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

Dans ce cas, plusieurs détails inexistantes réellement peuvent s'ajouter à la lésion. D'où, certaines régions normales du tissu mammaire peuvent être considérées comme lésions ce qui va engendrer des biopsies inutiles [39].

Les techniques traditionnelles d'amélioration du contraste ont été appliquées à la mammographie pendant plus de trois décennies. La méthode du masque flou a été utilisée par Chan et al. [43]. Des études concernant les caractéristiques opérationnelles du récepteur ont

montré que la méthode du masque flou parvient à améliorer la détection des microcalcifications. Cependant, cette méthode a causé l'augmentation du bruit dans l'image [39].

Une approche couramment utilisée pour l'amélioration du contraste est la modification globale de l'histogramme. Cette méthode consiste à réaffecter les valeurs d'intensité des pixels afin de rendre la nouvelle répartition des intensités plus uniforme. Ceci peut être réalisé par l'égalisation d'histogramme ou par l'étirement d'histogramme [44]. Toutefois, ces transformations ont l'inconvénient de faire ressortir le bruit dans les images déjà bruitées.

En tenant compte de ces résultats, on peut dire que les transformations globales d'amélioration du contraste sont des techniques non adaptées aux informations locales contenues dans l'image. Il existe une variabilité significative dans la taille et la forme des lésions mammaires et les techniques classiques fournissent souvent des résultats non satisfaisants [39].

Pour cette raison, il s'avère nécessaire de mettre au point des approches locales d'amélioration du contraste plus appropriées à la nature des images mammographiques. Dans ce contexte, plusieurs approches basées sur les informations statistiques locales et les corrélations d'ordre supérieur ([45] ; [46]) ont été proposées.

Une alternative proposée en tant que transformation locale est l'amélioration du contraste basée sur le voisinage adaptatif (Adaptive Neighborhood contrast Enhancement (ANCE)). Dans la méthode ANCE, chaque pixel est considéré comme un pixel germe (seed pixel) dans une procédure de croissance de région. La procédure de croissance de région identifie l'ensemble des pixels qui sont similaires et connectés au pixel germe [47]. Les auteurs dans [48] sont les premiers à signaler l'importance du traitement d'image basé sur le voisinage adaptatif pour améliorer le contraste des images mammographiques. Rangayyan and Nguyen [49] ont défini une nouvelle méthode de croissance de région qui pourrait avoir des formes arbitraires plutôt que des formes carrées.

D'autres méthodes ont été proposées et se sont basées sur les transformations d'ondelettes. Ces méthodes varient selon l'ondelette mère et les coefficients utilisés et fonctionnent comme un filtre passe-bande ce qui leur permet d'améliorer le contraste des différentes lésions indépendamment de leur taille et de leur type [50]. Sivaramakrishna et al. [51] ont comparé la performance de quatre algorithmes de rehaussement du contraste dans les deux cas d'images mammographiques contenant des microcalcifications et contenant des masses. Les algorithmes comparés sont : l'amélioration du contraste basée sur le voisinage adaptatif ANCE [47], le masque flou adaptatif [45], l'égalisation d'histogramme basée sur le contraste limité adaptatif [44] et les ondelettes [50]. Dans la majorité des cas contenant des

microcalcifications, l'algorithme ANCE fournit le meilleur résultat de rehaussement du contraste. Dans le cas d'images contenant des masses, on remarque que les images originelles sont plus nettes que celles rehaussées.

Certaines méthodes de rehaussement d'images mammographiques causent une amplification du bruit (la méthode du masque flou, la modification globale de l'histogramme) et par la suite la distorsion des lésions (qui présentent l'information la plus importante à préserver dans de telles images). Les radiologues préfèrent que les images rehaussées maintiennent la même apparence que les images originelles ce qui limite les compétences des techniques de rehaussement [39]. De plus, en se basant sur les résultats fournis par [51], le rehaussement du contraste dans notre cas de traitement des masses mammaires ne s'avère pas nécessaire.

2.3.2 Segmentation :

La segmentation est souvent considérée comme l'étape initiale dans un système de diagnostic assisté par ordinateur (DAOx) surtout si on fait abstraction de l'étape de prétraitement qui, d'après la section précédente, n'est pas indispensable dans le cas de traitement des masses. La phase de segmentation est très importante puisque les traitements ultérieurs (description et classification) sont fortement liés au résultat de segmentation. En effet, une bonne détection du contour de la lésion engendre une description fidèle à ses caractéristiques. Ainsi, on peut garantir une classification minimisant le taux des faux positifs et maximisant le taux des vrais négatifs [39].

Cette étape consiste à détecter ou bien la totalité du sein à partir du fond de l'image ou bien un type d'anomalie bien spécifique comme les microcalcifications et les masses. Il a été démontré que la détection des masses est plus difficile que la détection des microcalcifications. En effet, les masses peuvent être masquées partiellement par le tissu mammaire. Par ailleurs, il y a une grande variabilité dans ces lésions (forme, contour, dimension) [39].

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

une certaine homogénéité. Ces méthodes permettent d'aboutir directement à une partition de l'image étant donné que chaque pixel est affecté à une région unique. On détaille dans ce qui suit les trois approches citées précédemment [39].

2.3.2.1 Les méthodes de seuillage :

Le seuillage a pour objectif de segmenter une image en deux ou plusieurs classes. Cette opération consiste à effectuer une partition de l'histogramme en niveaux de gris en utilisant un ou plusieurs seuils. Chaque pic de l'histogramme correspond à une classe [39]. En effet, cette méthode n'est efficace que si l'histogramme contient réellement des pics séparés. Les méthodes de seuillage ont été largement utilisées pour la segmentation de masses mammaires. Par exemple, les auteurs dans ([53] ; [52]) ont utilisé différentes valeurs de seuils en niveau de gris qui dépendent du type de tissu mammaire et ceci en se basant sur une analyse de l'histogramme. Plus récemment, Mudigonda et al. [54] ont utilisé un seuillage multiniveau pour détecter des contours fermés. L'inconvénient majeur de cette approche est le fait de considérer que les masses ont une densité uniforme par rapport au fond de l'image ce qui n'est pas toujours vérifié.

Dans certains cas, le seuillage n'est pas appliqué directement à l'image mammographique. On fait appel d'abord à une étape de prétraitement pour améliorer la qualité de l'image [39]. Cependant, l'inconvénient majeur de ces différentes techniques de seuillage est le choix du seuil ou de l'intervalle de seuillage. En effet, avec un intervalle trop large, on obtient des faux positifs. Dans ce cas, l'image seuillée contient des pixels qui ne font pas partie des objets d'intérêt. Il s'agit généralement de bruit ou de pixels (zones) qui ont un niveau de gris proche de celui des objets recherchés. Avec un intervalle trop étroit, on obtient des faux négatifs. Certains objets d'intérêt ou bien des parties de ces objets n'apparaissent pas dans l'image seuillée [39].

2.3.2.2 Les méthodes basées régions :

Les méthodes basées région regroupent des pixels ayant des propriétés communes, ainsi ces méthodes donnent directement une partition de l'image. Les méthodes basées région les plus populaires sont la méthode de croissance de régions (region growing) ([67] ; [68] ; [69]) et la méthode de division et fusion de régions (split and merge) ([70] ; [71] ; [72]).

L'étape la plus critique dans les méthodes de segmentation par croissance de région est le choix des points germes de départ. En effet, l'étape de croissance est basée sur l'utilisation d'un critère d'homogénéité. Si le point germe est situé dans une zone non homogène, la mesure de similarité va produire de fortes variations et la croissance va s'arrêter très rapidement. Par

conséquent, il convient de choisir les points de départ dans les zones les plus homogènes possibles.

Cette approche n'a pas été suffisamment exploitée dans le cadre de la segmentation des masses mammaires. Rangayyan et al. [71] ont introduit la mesure d'actance. Pour cet effet, le contour des régions d'intérêts a été approximé itérativement en utilisant la méthode de "division et fusion".

2.3.2.3 Les méthodes basées contours :

Les approches les plus anciennes de détection de contour font intervenir la notion de dérivation afin de détecter les transitions entre les objets et le fond de l'image qui correspondent à des zones de fortes variations d'intensité. Les méthodes dérivatives utilisées se basent sur des opérateurs tels que Roberts [55], Sobel [56], Prewitt [58] et Canny [57]. Une fois la norme et la direction du gradient sont calculées en chaque pixel de l'image, ces méthodes extraient des contours d'un seul pixel d'épaisseur en sélectionnant les maximas locaux des normes des gradients.

En mammographie, on est souvent confronté à la présence de bruit (d'acquisition), de textures fines ou de frontières pas très nettes, d'où, les transitions détectées ne correspondent pas forcément aux contours réels. Il est alors souvent nécessaire d'appliquer un traitement en aval afin d'écarter les transitions dues aux bruits. De plus, les contours extraits sont généralement discontinus et peu précis. Il faut donc, utiliser des techniques de reconstruction de contours par interpolation ou connaître a priori la forme de l'objet recherché afin de connecter les points du contour [39].

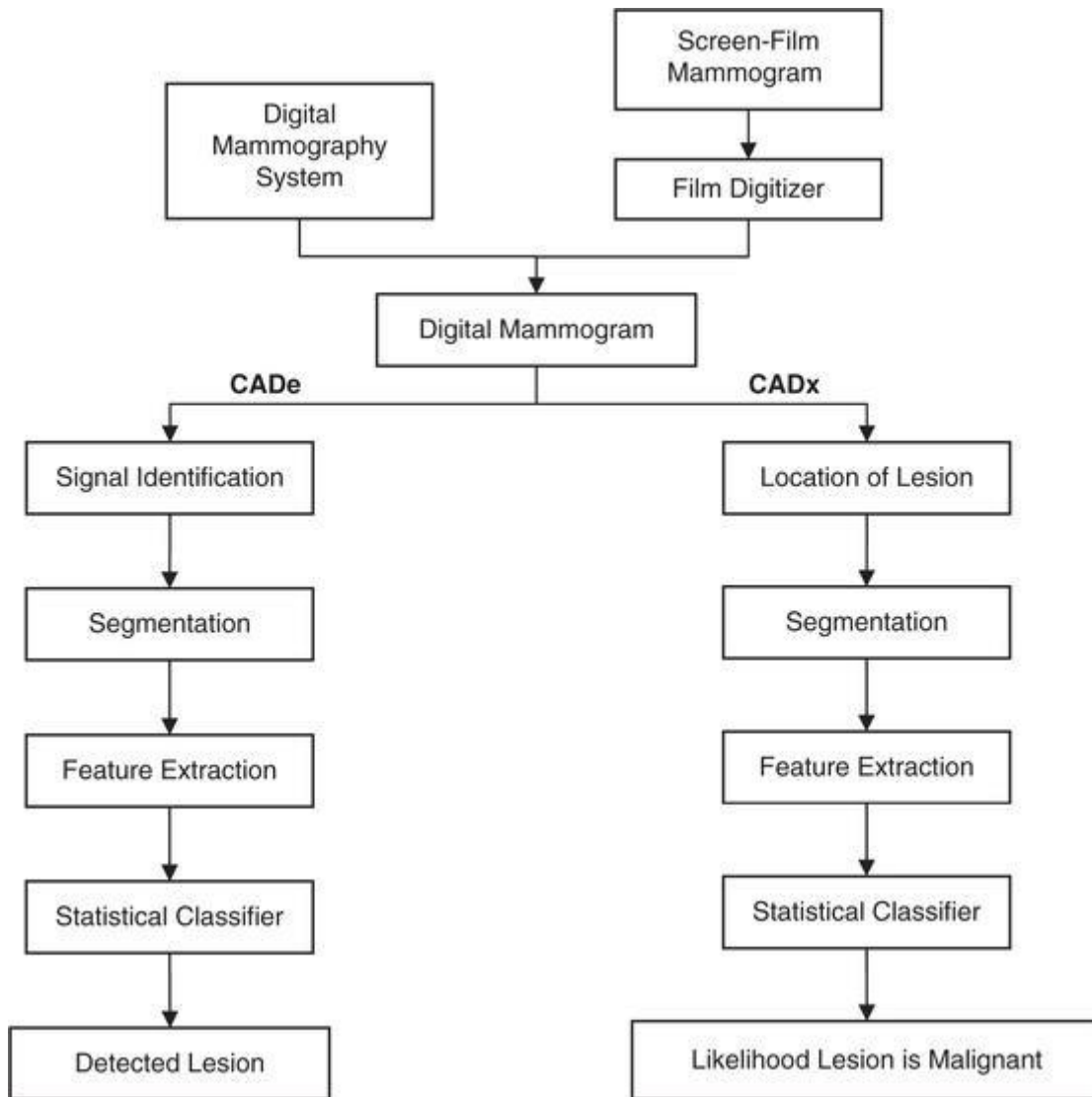


Figure.2.2. Les différents systèmes CAD (computer-aided diagnosis/Detection) [59]

2.3.3 Extraction des caractéristiques : Description

L'être humain reçoit en permanence des informations très diverses et très complexes par l'intermédiaire de ses cinq sens. En dépit de l'abondance de ces informations, le cerveau humain est capable de restituer chaque objet observé et de lui attribuer une représentation cohérente appelée *_description humaine_* [39].

Dans le domaine du traitement de l'image, la description est l'étape qui cherche à reproduire le même processus d'analyse et d'interprétation. En effet, la description a pour but d'extraire les caractéristiques qui décrivent au mieux et de façon quantitative ou qualitative les objets présents dans l'image. Elle transforme les informations de bas niveau issues de la phase d'acquisition (après probablement prétraitement et segmentation) en informations de haut niveau de telle sorte que les formes et les structures soient décrites de façon analytique. De

manière générale, plus la description effectuée est proche de _la description humaine_, plus elle est considérée comme robuste et fidèle à l'image initiale. Les méthodes de description d'images sont variables et dépendent de l'objectif visé (description globale, description locale) et du type d'image à analyser (image binaire, image en niveau de gris, image couleur). En littérature, la description d'images est assurée en utilisant la couleur, la texture et/ou la forme. Les différents types de descriptions et leur utilité sont détaillés dans les sections qui suivent [39].

2.3.3.1 La couleur :

En traitement d'images, une grande partie des images numériques à analyser est de type couleur. La notion de couleur est la perception subjective qu'a l'œil d'une ou plusieurs fréquences d'ondes lumineuses. On appelle espace de couleurs la représentation mathématique d'un ensemble de couleurs. Il en existe plusieurs, parmi lesquels les plus connus sont le RVB (Rouge, Vert, Bleu), le TSL (Teinte, Saturation, Luminance), le CMJ (Cyan, Magenta, Jaune), le CMJN (Cyan, Magenta, Jaune, Noir pur) et le CIE (codage de la couleur selon la Commission Internationale de l'Eclairage) [39]. Afin de décrire les différentes images couleurs, plusieurs attributs couleurs ont été proposés dans le domaine de l'imagerie notamment dans les systèmes de reconnaissance d'images par le contenu [85] les images satellitaires et la reconnaissance d'objets couleurs. Les attributs couleur les plus utilisés sont la caractéristique couleur moyenne, la caractéristique histogramme couleur, l'indexation couleur constante, les fonctions des moments de distribution couleurs et la caractéristique corrélogramme [39].

Dans le domaine du cancer du sein, l'étape de description des pathologies mammaires est un maillon très important dans les systèmes d'aide au diagnostic. En effet, le résultat de classification est étroitement lié au résultat fourni par cette étape. En dépit de la robustesse du classifieur utilisé, on ne peut garantir un bon résultat de classification que si la description de l'anomalie a été accomplie avec le maximum de précision. Le rôle principal de cette étape est de fournir aux radiologues des informations précises concernant les aspects pathologiques des lésions mammaires et de spécifier la gravité de la tumeur en termes de malignité/bénignité. Toutefois, les images mammographiques du sein sont des images en niveau de gris qui ne contiennent pas d'information couleur. De ce fait, ce type de description d'images ne sera pas pris en compte dans notre étude.

2.3.3.2 La texture :

Dans le traitement d'image, l'analyse de texture consiste à calculer une série de mesures dans le but de définir une texture perçue sur une image. L'analyse de texture renvoie des

informations sur l'arrangement spatial des couleurs ou des intensités dans tout ou partie de cette image

Dans le domaine du traitement de l'image, il n'existe pas de définition conventionnelle de la texture. Les définitions mathématiques proposées peuvent être trop générales et donc imprécises et peuvent aussi être trop détaillées et donc ne couvrent pas tous les cas de texture envisageables [39].

2.4.3.2.1 Les statistiques du premier ordre :

Les statistiques du premier ordre mesurent la probabilité d'observer un niveau de gris à un emplacement aléatoire d'une image. Elles sont estimées sur les niveaux de gris sans la considération de leur distribution relative. Les statistiques du premier ordre sont calculées à partir de l'histogramme des intensités de l'image noté 'Hist'. Ce vecteur est obtenu en utilisant la fonction suivante [60] :

$$\text{Hist}(i) = \sum p \text{ tel que } I(p) = i \quad \text{Équation 1}$$

Avec 'I(p)' est l'intensité du pixel 'P' dans l'image, $i \in \{1, 2, 3, \dots, ng\}$, tel que 'ng' est le nombre de niveaux de gris.

Parmi les statistiques du premier ordre, nous citons la moyenne (Moy), l'écart type (σ_{Hist}) et la variance (Var) [59].

La moyenne :

$$\text{Moy} = \frac{1}{N} \sum_i^{ng} i \times \text{Hist}(i) \quad \text{Équation 2}$$

Où 'N' est le nombre de pixels dans l'image [72].

L'écart-type :

$$\sigma_{Hist} = \sqrt{\frac{1}{ng} \sum_{i=1}^{ng} (\text{Hist}(i) - \text{Moy})^2} \quad \text{Équation 3}$$

La variance :

$$\text{Var} = \frac{1}{N} \sum_{i=1}^p (I(p) - \text{Moy})^2 \quad \text{Équation 4}$$

L'Asymétrie :

$$\text{Asym} = \frac{\sum_{i=0}^{n-1} P_i (x_i - \text{Moy})^3}{\sum_{i=0}^{n-1} P_i (x_i - \text{Moy})^{3/2}} \quad \text{Équation 5}$$

Aplatissement :

$$\text{Aplt} = \frac{1}{N} \frac{\sum_i \sum_j (P(i, j) - \text{Moy})^4}{\text{Var}^4} \quad \text{Équation 6}$$

Entropie :

$$\text{Ent} = - \sum_i \sum_j P(i, j) \log P(i, j) \quad \text{Équation 7}$$

Uniformité :

$$Unif = \sum_i P_i^2(z_i) \quad \text{Équation 8}$$

Régularité :

$$Reg = 1 - \frac{1}{1 - \sigma^2(z)} \quad \text{Équation 9}$$

Où : $\mathbf{P}(\mathbf{i}, \mathbf{j})$ est l'intensité du pixel à la position (\mathbf{i}, \mathbf{j}) . \mathbf{N} représente le nombre total des pixels

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

L'inconvénient majeur des statistiques du premier ordre basées sur l'histogramme est qu'elles n'exploitent que l'information niveau de gris et ne prennent pas en considération les relations spatiales entre les pixels. La description de texture à base des statistiques d'ordre supérieur (statistiques du second ordre par exemple) s'avère plus efficace [60].

2.3.3.2.2 La matrice de co-occurrence :

Les méthodes statistiques du second ordre sont les méthodes qui mettent en jeu deux pixels pour le calcul des caractéristiques [60]. La principale méthode du second ordre utilisée pour le traitement des images médicales notamment les images mammographiques est la méthode de matrice de co-occurrence ou méthode de dépendance spatiale des niveaux de gris : SGLD (Spatial Gray Level Dependency) développée par Haralick [61]. C'est une méthode statistique qui consiste à construire des matrices de co-occurrence pour représenter les relations entre les pixels d'une image.

La matrice représente la probabilité conjointe pour deux niveaux de gris i, j soient dans une relation spatiale donnée. Cette relation est définie en termes de la distance et d'angle entre ces deux pixels. L'angle permet d'évaluer la direction de texture et l'application de plusieurs valeurs de distance peut donner une description significative de la taille de la périodicité de la texture. Le choix de la distance et de la combinaison de l'angle, ainsi que le niveau de quantification, est quelque peu arbitraire [60].

Les matrices de cooccurrences à niveaux de gris (GLCM), introduites par Haralick et al.

[61], sont très intéressantes car elles mesurent la distribution des niveaux de gris dans l'image tout en prenant en compte les dépendances spatiales entre les pixels. Cette matrice est une matrice carrée de taille n^2 où n représente le nombre des niveaux de gris de l'image. La plupart des images sont codées sur 256 niveaux de gris par conséquent, la taille des GLCM est considérable. Dans la pratique ces niveaux de gris sont requantifiés (16, 32 ou 64 niveaux de gris) ce qui permet d'avoir un temps de traitement plus restreint [Paulhac 2009], dans notre cas les images sont codées sur 4096 niveaux de gris et nous avons utilisés 64 niveaux de gris pour les GLCM [60].

Une matrice de cooccurrence mesure la probabilité d'apparition des paires de valeurs de pixels (P_i, P_j) situés à une certaine distance d dans l'image. Elle est basée sur le calcul de la probabilité $P(i, j, d, \theta)$ qui représente le nombre de fois où un pixel de niveau de gris i apparaît à une distance relative d d'un pixel de niveau de gris j et selon une orientation θ donnée. Afin de ne prendre en compte qu'une information très locale de voisinage, la distance d n'excède pas en générale quelques pixels. Les directions angulaires θ classiquement utilisées sont $0, \pi/4, \pi/2$ et $3\pi/4$. Fig. 2.5 illustre ce propos [60].

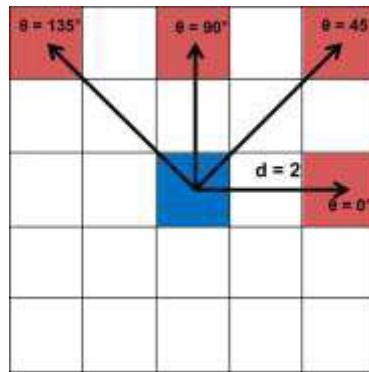


Figure 2.5: Les voisins d'une distance $d = 2$ du pixel 'x' selon les 4 directions.

Dans cette méthode, il faut calculer la matrice de co-occurrence pour une région d'intérêt (ROI) carré à quatre angles différents : $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ et à une distance d . En conséquence, les quatre matrices de co-occurrence sont générés pour chaque valeur de distance d , ces matrices sont additionnées pour obtenir une matrice de rotation invariant M^d :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad \text{Équation 5}$$

Comme les matrices de cooccurrences incluent beaucoup d'informations et sont donc consommatrices en espace mémoire, elles ne sont pas directement utilisées pour caractérisation des textures. Les utilisateurs préfèrent donc calculer à partir de ces matrices quelques descripteurs afin de réduire la quantité d'informations à analysés, tout en conservant la pertinence de ces caractéristiques. Considérons l'image (Figure 2.6: exemple d'une image.) [60].

1	3	2	5	4	3
2	3	4	2	3	4
3	4	5	6	2	5
7	8	5	6	4	2
3	4	5	2	4	6

Figure 2.6 : exemple d'une image.

Si la direction θ est nulle (direction horizontale) et la distance $d=1$, alors la matrice de cooccurrence

$M_{i,j}^d$ Est calculée comme suit :

$Ng \backslash Ng$	1	2	3	4	5	6	7	8
1	0	0	1	0	0	0	0	0
2	0	0	2	1	2	0	0	0
3	0	1	0	4	0	0	0	0
4	0	2	1	0	2	1	0	0
5	0	1	0	1	0	2	0	0
6	0	1	0	1	0	0	0	0
7	0	0	0	0	0	0	0	1
8	0	0	0	0	1	0	0	0

Figure 2.7: La matrice de cooccurrence $M_{i,j}^d$ de l'image (Fig. 2.6).

Haralick et Shanmugam [61] dans leur article "Textural features for image classification", proposent quatorze descripteurs texturaux extraits à partir des matrices de cooccurrences.

Le calcul de ces descripteurs est basé sur les matrices de cooccurrence. Ces matrices qui mettent effectivement en avant les relations existantes entre les pixels de l'image en faisant intervenir l'aspect local (les niveaux de gris) et l'aspect spatial (d, θ). Ces descripteurs sont les suivants

1. L'énergie (uniformité) :

Ce descripteur mesure l'uniformité de la texture. Il atteint de grandes valeurs lorsque la distribution des niveaux de gris est constante ou de forme périodique. Dans ce dernier cas, lorsque (d, θ) correspond à la période, les valeurs élevées d'énergie sont obtenues pour les matrices M_{θ}^d [60].

$$Energie^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad \text{Équation 6}$$

Où : G est le nombre de niveaux de gris utilisé.

2. Le contraste :

Le contraste mesure les variations locales des niveaux de gris, lorsque ces variations sont importantes, alors le contraste sera élevé. La valeur en est d'autant plus élevée que la texture présente un fort contraste. Ce paramètre est fortement non corrélé à l'énergie [59].

$$Contraste^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 \times M_{i,j}^d \quad \text{Équation 7}$$

3 La corrélation :

La corrélation permet de déterminer si certaines colonnes de la matrice sont égales, c'est-à-dire s'il existe des dépendances linéaires dans l'image. En effet, plus les valeurs sont

uniformément distribuées dans la matrice de cooccurrences et plus la corrélation est importante. La corrélation n'est corrélée ni à l'énergie, ni à l'entropie [60].

$$\text{Corrélation}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i-u_i^d)(j-u_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad \text{Équation 8}$$

Où : u_i^d, σ_i^d : sont la moyenne et l'écart-type de la ligne i de la matrice $M_{i,j}^d$

u_j^d, σ_j^d : sont la moyenne et l'écart-type de la colonne j de la matrice $M_{i,j}^d$

1. *L'homogénéité (Moment différentiel inverse) :*

L'homogénéité représente le comportement inverse de celui du contraste. En effet, plus la texture possède de régions homogènes et plus le moment différentiel inverse est élevé [59].

$$\text{Homogénéité}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1+(i-j)^2} \right] \quad \text{Équation 9}$$

2. *La variance :*

La variance représente l'hétérogénéité de la texture. Elle augmente lorsque les niveaux de gris diffèrent de leur moyenne. Cette mesure est indépendante du contraste. La variance correspond au carré de l'écart-type :

$$\text{Variance}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - u^d)^2 \quad \text{Équation 10}$$

Où : μ est la valeur moyenne de la matrice de rotation invariante $M_{i,j}^d$ [60].

6. *La somme moyenne :*

$$SM^d = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad \text{Équation 11}$$

7. *L'entropie de la somme :*

$$SM^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j) M_{i,j}^d \log_{(i+j)M_{i,j}^d} \quad \text{Équation 12}$$

8. *La variance de la somme :*

$$VS^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [(i + j) - ES^d]^2 M_{i,j}^d \quad \text{Équation 13}$$

9. *Entropie :*

L'entropie caractérise la complexité de l'image. Lorsque les valeurs de la matrice de cooccurrences sont égales, l'entropie prend une valeur élevée. Elle permet ainsi de représenter le degré de granulation de l'image. En effet, plus la valeur de l'entropie est grande et plus la granulation est grossière. L'entropie est faible si l'image possède souvent les mêmes couples de pixels. C'est un indicateur de désordre. L'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire (sans structure apparente). Cette mesure est fortement corrélée (par l'inverse) à l'énergie [60].

$$\text{Entropie}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \log_{M_{i,j}^d} \quad \text{Équation 14}$$

10. Entropie de la différence :

$$ED^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j) M_{i,j}^d \log_{(i-j)M_{i,j}^d} \quad \text{Équation 15}$$

11. Variance de la différence :

$$VD^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [(i-j) - ED^d]^2 M_{i,j}^d \quad \text{Équation 16}$$

12. Information sur la corrélation :

$$\text{Corr1} = \frac{(\text{Entropie}^d - HXY1)}{\text{Max}(HX, HY)} \quad \text{Équation 17}$$

Où HX et HY sont respectivement les entropies de M_i^d et M_j^d

$$HXY1 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \log_{((i \times j) M_{i,j}^d)} \quad \text{Équation 18}$$

13. Information sur la corrélation 2 :

$$\text{Corr2} = (1 - \exp^{-2(HXY2 - \text{Entropie}^d)})^{1/2} \quad \text{Équation 19}$$

Où

$$HXY2 = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i \times j) M_{i,j}^d \log_{((i \times j) M_{i,j}^d)} \quad \text{Équation 20}$$

14. Coefficient de corrélation maximal :

$$\text{CorrMax} = (2^{\text{eme Plus grande valeur propre de } Q})^{1/2} \quad \text{Équation 21}$$

$$Q(i, j) = \frac{\sum_{k=0}^{G-1} (M_{i,k}^d \times M_{j,k}^d)}{((i \times j) M_{i,j}^d)} \quad \text{Équation 22}$$

Les descripteurs 11, 12, 13, 15, 16, 17, 19 et 21 apportent des informations supplémentaires sur la corrélation et les degrés d'homogénéité et de complexité de l'image [60].

2.3.3.2.3 Local Binary Pattern (LBP) méthode des motifs locaux binaires :

Les motifs locaux binaires (Local Binary Patterns en anglais) sont des caractéristiques utilisés en vision par ordinateur pour reconnaître des textures ou pour la détection d'objets dans des images numériques.

$$e_p = \begin{cases} 0, & \text{si } (w_p > w_c) \\ 1, & \text{si } (w_p = w_c) \\ 2, & \text{si } (w_p < w_c) \end{cases} \quad \text{Équation 1}$$

Où p et e est calculé avec un nombre de niveaux $L=3$, avec $p=1, \dots, 8$ en comparant la valeur d'intensité d'un centre pixel w_c avec un pixel voisin w_p dans le sens des aiguilles d'une montre ou dans le sens inverse des aiguilles d'une montre. Le nombre de fonctionnalités peut être défini par $3^8 = 6561$. Ensuite, la texture d'une image peut être décrite avec un histogramme de ces valeurs.

Plus tard en 1994, les niveaux d'opérateur ont été réduits de $L = 3$ à $L = 2$, et les comparaisons ont été effectuées par booléenne opération en premier opérateur LBP [6]. Les

valeurs d'intensité ont été calculées en comparant le pixel de voisinage avec valeur de pixel centrale, se déplaçant de haut en bas et de gauche à droite comme dans l'expression suivante [73] :

$$e_p = \begin{cases} 1, & (w_p > w_c) \\ 0, & \text{Autrement} \end{cases} \quad \text{Équation 2}$$

Avec $L = 2$, la taille du vecteur de caractéristiques est réduite et le rend compactable avec des valeurs de 2^8 . Cependant, cette compacité a un coût : ces descripteurs sont plus sensibles au bruit, en particulier dans les régions quasi uniformes d'une image [66,64].

LBP proposé par Ojala et. Al. [65], il est très célèbre en raison de sa représentation simple, de son calcul efficace, de sa faible complexité de calcul et capacité de codage des détails fins. Le descripteur LBP est la combinaison de statistiques et modèle basé sur la structure. Les autres extensions très puissantes de LBP avec invariance de rotation multi-résolution également ont été développés par Ojala et. Al. [66].

L'opérateur LBP d'origine est invariant par rotation à la texture en échelle de gris [64]. L'extraction de l'opérateur LBP comprend deux étapes principales : la première est le seuillage et la seconde est l'étape d'encodage. L'étape de seuillage extrait les informations sur la différence binaire locale (1 ou 0) en comparant la valeur des pixels voisins circulaires avec le pixel central dans un patch d'image [73].

$$S(x) = (w_c, w_0 - w_c, w_1 - w_c, \dots, w_{p-1} - w_c) \quad \text{Équation 3}$$

Où w_c et w_p sont définis dans l'équation (1) et le nombre d'échantillonnage p ($p= 0, \dots, p-1$) dans l'équation (2) [73].

$$S(x) = \begin{cases} 1, & \text{si } x \geq 1 \\ 0, & \text{Autrement} \end{cases} \quad \text{Équation 4}$$

Ensuite, dans la deuxième étape de codage, les nombres binaires obtenus à partir de la première étape sont convertis en nombre décimal comme illustré à la figure 2.8.

La première version de l'opérateur LBP fonctionne avec un patch 3×3 d'une image, mais plus tard, cette taille de patch n'a pas de règle fixe pour en utilisant des points d'échantillonnage dans le cercle dans une autre version d'extension [76]. Au lieu d'utiliser la taille 3×3 du patch, il peut être changer avec les valeurs de P, R telles que (8,1), (16,2), (8,2) et ainsi de suite. L'équation mathématique peut être définie comme suivre [73] :

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(w_p - w_c) \times 2^p \quad \text{Équation 5}$$

Où w_c, w_p et p sont déjà définis dans les équations (1), (2) et (3). Et P, R est le nombre de voisins ou respectivement les échantillons et le rayon du cercle [73].

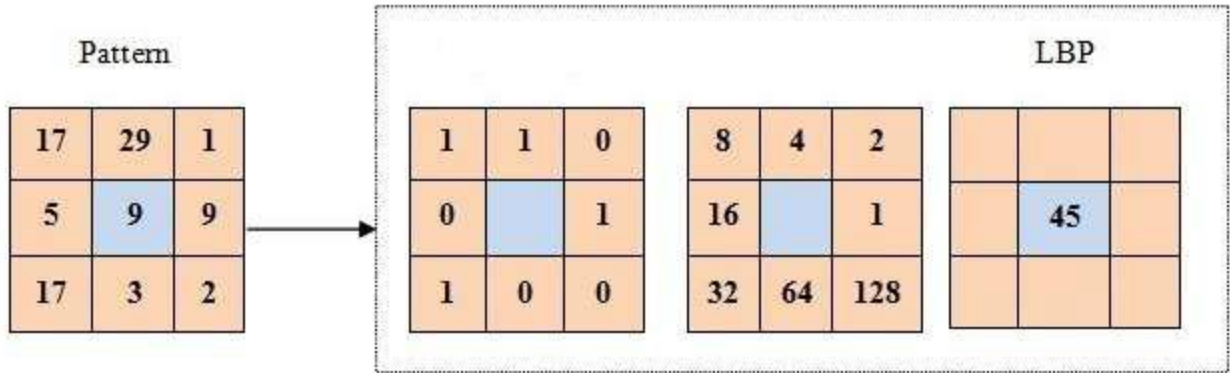


Figure 2.8 : Calcul LBP sur patch 3×3 sur image

Une autre variation de l'opérateur d'origine semble sous la forme de modèles uniformes [62]. Un motif uniforme peut être indiqué avec « u » et défini comme son uniformité (nombre de transactions de 0 à 1 ou de 1 à 0 bitwise) est tout au plus 2 [73]. Il y a les principales raisons de considérer uniquement le modèle uniforme, qui dans le scénario réel, la majeure partie de la texture ont de nature uniforme et eux-mêmes sont plus stables, c'est-à-dire moins affectés par le bruit et une autre raison, car considérez uniquement le modèle uniforme, de sorte que les échantillons minimaux de LBP sont fiables pour reconnaître le modèle de texture [63].

2.3.3.3 La forme :

De même que les notions de couleur et de texture, la notion de forme est très importante puisqu'elle nous permet d'identifier les objets qui nous entourent. L'analyse de forme n'est considérée réussie que si elle permet de décrire les objets de façon similaire à la perception humaine des formes. Toutefois, il n'existe pas de définition conventionnelle pour caractériser cette notion. La compréhension de la notion de forme nécessite de définir certaines notions de topologie telles que la distance, l'ouverture, ... [74]. **Distance :** Pour le calcul de distance, on utilise généralement la norme euclidienne. La métrique euclidienne définit la distance entre deux points comme la longueur du segment les reliant. Pour tous points $p(x_p, y_p)$ et $q(x_q, y_q)$ de \mathbb{R}^2 , la distance euclidienne est l'application $d_E(p, q) : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par :

$$d_E(p, q) = \|p - q\| = \sqrt{(x_q - x_p)^2 + (y_q - y_p)^2}$$

(2.1) La distance est une application vérifiant la symétrie ($d_E(p, q) = d_E(q, p)$), la séparation ($d_E(p, q) = 0 \Rightarrow p \equiv q$) et l'inégalité triangulaire ($d_E(p, q) \leq d_E(p, r) + d_E(r, q)$). Dans le cas où Ω est un ensemble de points de \mathbb{R}^2 et p un point de \mathbb{R}^2 alors la distance euclidienne entre Ω et p notée par $d_E(p, \Omega)$ est désignée par la plus courte distance entre p et Ω :

$$d_E(p, \Omega) = d_E(\Omega, p) = \min_{q \in \Omega} d_E(p, q) \quad (2.2)$$

Boule ouverte : On désigne par p un point de \mathbb{R}^2 , r un réel positif et d_E une distance. On appelle boule ouverte $B^{d_E}(p, r)$ de centre p et de rayon r , l'ensemble des points x tel que la distance $d_E(p, x)$ est strictement inférieure à r :

$$B^{d_E}(p, r) = \{x \in \mathbb{R}^2 \mid d_E(p, x) < r\} \quad (2.3)$$

Ouvert : En mathématiques, et plus particulièrement en topologie, on appelle ensemble ouvert ou partie ouverte ou tout simplement un ouvert, tous ensemble Ω de \mathbb{R}^2 qui ne contient aucun point de sa frontière. En d'autres termes, si pour tout point p de Ω il existe un réel positif r tel que la boule ouverte $B^{d_E}(p, r)$ soit entièrement contenue dans Ω .

Fermé : Le complémentaire d'un ensemble Ω dans \mathbb{R}^2 , noté Ω^c est l'ensemble des points p qui n'appartiennent pas à Ω . Un ensemble de \mathbb{R}^2 est dit fermé si son complémentaire est un ouvert. On appelle fermeture d'un ensemble de points Ω : L'ensemble $\bar{\Omega}$ défini par l'intersection des fermés contenant Ω .

Borné : Un ensemble de points Ω de \mathbb{R}^2 est dit borné s'il existe un point p de \mathbb{R}^2 et un réel positif r de façon que Ω soit un sous-ensemble de la boule ouverte $B^{d_E}(p, r)$.

Intérieur : On appelle intérieur $Int(\Omega)$ d'un ensemble de points Ω de \mathbb{R}^2 le plus grand ouvert de Ω , qui est en effet la réunion de tous les ouverts inclus dans Ω .

Contour : La notion de contour est fortement liée à la notion de forme. Ainsi, l'introduction de cette notion de contour facilite la compréhension de la notion de forme. Pour ceci, on considère un ensemble de points Ω de \mathbb{R}^2 , le contour C de Ω est défini par l'intersection entre la fermeture de Ω et la fermeture de son complémentaire ($C = \bar{\Omega} \cap \bar{\Omega}^c$).

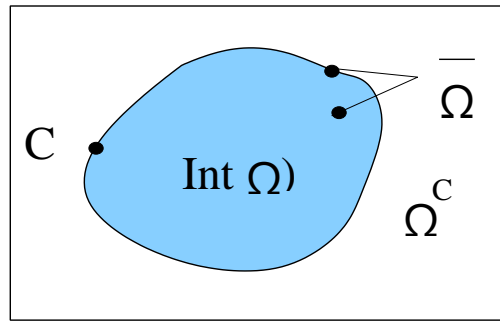


Figure 2.9–Schéma détaillant les notions de : fermeture, intérieur, contour et complémentarité [39].

Afin de rendre les notions d'intérieur, de fermeture, de complémentaire et de contour plus accessible, on propose la figure 2.3. Dans cette figure, l'intérieur $int(\Omega)$ de l'objet est désigné par la partie coloriée, le contour C est la courbe fermée noire alors que la fermeture $\bar{\Omega}$ est représentée par l'ensemble (partie coloriée + contour noir). Le complémentaire de l'intérieur $int(\Omega)^c$ est l'ensemble (contour + reste du plan), le complémentaire de la fermeture Ω est le reste du plan sans considérer le contour alors que le complémentaire du contour C^c est constitué par l'ensemble (reste du plan + intérieur) [39].

Forme : En se basant sur les notions définies ci-dessus, la forme peut être définie en tant qu'un ensemble de points Ω de \mathbb{R}^2 qui est la fermeture d'un sous-ensemble de \mathbb{R}^2 ouvert et borné. Le contour de Ω est représenté par un nombre fini de courbes fermées. Une fois la notion de description (de couleur, de texture et de forme) est définie d'une manière générale, reste à détailler l'étape de description analytique des lésions mammaires par le biais de descripteurs spécifiques. Une telle description permet de caractériser et d'analyser la forme de façon à rendre l'information accessible aux systèmes informatiques dédiés au traitement automatique des mammographies. On clôture ce chapitre par la classification qui représente la dernière étape d'un système de diagnostic assisté par ordinateur [39].

2.3.4 Classification :

La classification est considérée comme la dernière étape dans un système de diagnostic assisté par ordinateur (DAOx). Elle exploite le résultat de description (qui lui-même exploite le résultat de segmentation) pour pouvoir décider de la nature pathologique de la masse [39].

La notion de classification signifie l'affectation d'une étiquette à des échantillons d'une base de données en utilisant un certain nombre de caractéristiques. Ces caractéristiques doivent bien évidemment être capables d'identifier chaque échantillon. En traitement d'images, l'échantillon peut désigner un pixel, une zone dans l'image, un objet représenté dans l'image

ou l'image elle-même. Selon l'application, le but de la classification est soit de : classifier les pixels de l'image en différentes zones. Dans ce cas, le problème de classification revient à un problème de segmentation d'images en différents objets. A titre d'exemple, on peut classifier les différentes zones d'une image mammographique en lésion ou non lésion.

classifier l'image ou les objets de l'image selon différentes catégories. A titre d'exemple, on peut classifier les masses qui se trouvent dans les images mammographiques en malignes ou bénignes. On peut distinguer deux catégories de méthodes de classification : les classifications non-supervisées et celles supervisées [39].

2.3.4.1 Les méthodes de classifications non supervisées :

Ces techniques sont utilisées lorsque l'identité des classes n'est pas connue. Cela résulte d'un manque d'information de la population à étudier. Il existe des algorithmes de classification, composés de plusieurs itérations, permettant de créer des regroupements d'individus ayant des caractères similaires. La classification non-supervisée, dite automatique, ou groupement connue en anglais par clustering consiste à déterminer les différentes classes naturellement sans aucune connaissance préalable. L'objectif, dans ce cas, est d'identifier une structure dans les images de la base en se basant sur leurs contenus. Les images sont attribuées aux différentes classes estimées selon deux critères essentiels qui sont la grande homogénéité de chaque classe et la bonne séparation entre les classes [39].

Parmi les méthodes de classification non supervisées la méthode la plus communément utilisée est celle de l'algorithme K-moyennes également appelée algorithme des nuées dynamiques (en anglais k-means) [75]. L'algorithme fonctionne en précisant le nombre K de classes (clusters) attendues (K étant fixe par l'utilisateur). Il calcule la distance intra-classe et refixe les centres de classe selon les valeurs de distance. Les inconvénients de cette méthode sont premièrement la nécessité de fixer le nombre de classes avant de commencer la classification. Deuxièmement, cette méthode est très sensible à la répartition initiale des données. Finalement, cette méthode suppose que les classes suivent des lois de distribution normales réduites, autrement dit, avec la même importance dans toutes les directions de l'espace ce qui n'est pas toujours vérifié [39].

Une autre méthode de classification non supervisée est la carte auto-organisatrice connu sous le nom anglais Self Organizing Map (SOM) [76]. Une SOM est un réseau de neurones qui, par un processus non-supervisé compétitif, est capable de projeter des données de grandes dimensions dans un espace à deux dimensions. Au cours de l'apprentissage, chaque neurone se spécialise dans la reconnaissance d'un certain type d'entrée. La carte auto-organisatrice est

composée d'un ensemble de neurones connectés entre eux. Une configuration entre l'espace d'entrée et l'espace du réseau est construite, ainsi, deux observations proches dans l'espace d'entrée activent deux unités proches sur la carte. Cette méthode est plus robuste aux conditions initiales que l'algorithme des K-moyennes. L'inconvénient majeur de cette méthode est le temps de calcul associé aux itérations qui permettent la construction de la carte auto-organisatrice [39].

2.3.4.2 Les méthodes de classification supervisées :

Si l'utilisateur possède suffisamment d'informations sur la population à étudier (tel est le cas des images mammaires), il peut effectuer une classification supervisée. Cette catégorie suppose avoir un groupe d'individus de chaque classe, dont on connaît leur appartenance. Ces individus forment des échantillons *_d'apprentissage_*. Ils sont utilisés pour entraîner le classifieur. D'autres échantillons, dits *_de test_*, servent à valider la classification en évaluant sa pertinence à travers le taux d'individus bien classés. Il existe plusieurs méthodes de classification supervisées. Les méthodes les plus réputées sont l'analyse discriminante linéaire, la régression logistique, les réseaux de neurones... [39] Certaines recherches se sont focalisées sur l'analyse discriminante linéaire. Il s'agit d'une méthode de classification simple qui sépare les images appartenant aux différentes classes en se basant sur une analyse linéaire. L'idée principale de cette technique est de construire les limites de décision directement en optimisant le critère d'erreur. Cependant, cette méthode n'est adaptée qu'aux données linéairement séparables ce qui n'est pas toujours le cas [39]. Les réseaux de neurones artificiels, connus par l'acronyme anglais ANN (Artificial Neural Networks), sont largement utilisés pour les problèmes de classification. Ils reposent sur la théorie des perceptrons. Un ANN est composé de plusieurs neurones repartis sur une couche d'entrée (désignant les descripteurs), une couche de sortie (désignant le résultat de classification) et un nombre de couches cachées. Par ailleurs, cette méthode est capable de modéliser des systèmes non linéaires très complexes. Toutefois,

l'inconvénient de cette méthode est le choix du nombre de couches cachées et du nombre de neurones dans chaque couche. Ainsi, l'utilisateur est amené à faire des essais avec différentes combinaisons du nombre de couches et de neurones afin d'aboutir au réseau de neurones le plus adapté à son type d'application [39].

Par contre, les réseaux de neurones à fonctions de base radiales, connus par l'acronyme anglais Radial Basis Functions (RBF), sont constitués par une seule couche cachée. L'avantage majeur par rapport aux autres réseaux de neurones artificiels est l'utilisation d'une structure moins complexe (une seule couche cachée). En plus, la complexité de calcul induite par leur

apprentissage est inférieure à celle induite par l'apprentissage des ANN grâce à l'existence d'algorithmes hybrides. Cependant, les performances d'un tel réseau dépendent, pour un choix de fonctions de base, du nombre de fonctions constituant la base de fonctions radiales (nombre d'unités de la couche cachée) et de l'estimation des paramètres du réseau [39].

D'autres recherches se sont orientées vers la régression logistique (RL). Il s'agit d'un modèle multivarié couramment utilisé en épidémiologie (ou oncologie). Elle s'utilise lorsque la variable de sortie (les classes) est qualitative, le plus souvent binaire (la survenue ou non d'une maladie). Les variables d'entrées (les descripteurs) peuvent être par contre soit qualitatives, soit quantitatives. La régression logistique est capable de réaliser une estimation de probabilité en utilisant une formulation logistique [39]. Face aux fonctions linéaires, les séparateurs à vaste marge, connues par l'acronyme anglais SVMs (Support Vector Machines), sont initialement conçues pour les problèmes de classification binaires. Elles permettent de séparer linéairement les exemples positifs des exemples négatifs dans l'ensemble des images d'apprentissage par un hyper-plan qui garantisse un maximum de marge [77]. L'efficacité des SVMs est souvent supérieure à celle de toutes les autres méthodes de classification supervisées. Pour les problèmes de non-séparabilité, les SVMs permettent d'effectuer une transformation non-linéaire des observations d'entrée dans un espace de dimension supérieure pour se ramener au cas linéaire. En outre, les SVMs peuvent s'adresser également aux problèmes de classification multi-classes. Trois annexes (Annexe A, B et C) sont fournies par la suite afin de détailler les méthodes de classification supervisées (RBF, RL et SVM) [39].

2.5 Quelques travaux sur les systèmes d'aide au diagnostic dédiés aux images mammographiques :

En 2020, V. S. Gnanasekaran et d'autres ont proposé un modèle de classification des masses mammaires à l'aide de la méthode CAD. MIAS, les ensembles de données auto-collectés et les ensembles de données DDSM sont utilisés dans ce étude. Prétraitement, segmentation, collecte et regroupement de fonctions sont utilisées dans ce système. Le système CAO comprend un modèle CNN composé de huit Co évolutionnaires, quatre max-pooling et deux couches entièrement connectées. Les résultats qu'ils obtiennent sont par la suite comparés aux filets pré-entraînés, Alex Net et VGG16, démontrant que le projet CNN a atteint une précision et une AUC plus élevées que ces deux des modèles. Le modèle proposé atteint des précisions de 92,54 %, 96,47 % et 95 % et des scores d'ASC de 0,85, 0,96, et 0,94 pour MIAS, DDSM et l'ensemble de données auto-collecté, respectivement. Une stratégie d'apprentissage extrême a

été utilisée pour cartographier la fusion de caractéristiques et extraire les caractéristiques CNN pour le sein détection et classification des cancers [92].

En 2020, V. S. Gnanasekaran et d'autres ont proposé une méthode pour détecter le cancer du sein à partir des mammographies. Prétraitement, segmentation, l'extraction des caractéristiques et la classification sont utilisées dans cette recherche. Ensuite, l'image est analysée, puis la segmentation est ajoutée. En effet les caractéristiques sont dérivées, et, troisièmement, la classification est effectuée. Après, une distinction est faite entre les différentes méthodes de classement. Machine à vecteurs de support (SVM), AdaBoost, arbre de décision, régression logistique, K plus proche voisin, et des classificateurs forestiers aléatoires sont utilisés pour classer le sein cancer. La précision obtenue est de 90%, 57%, 54%, 85%, 76 % et 61 % pour la machine à vecteurs de support (SVM), Ada-Boost, arbre de décision, régression logistique, K plus proche voisin, et des classificateurs forestiers aléatoires, respectivement, ce qui signifie que SVM atteint la plus grande précision de tous [92].

En 2021, B. Swiderski et autres ont proposé une méthode pour classer tumeurs du cancer du sein utilisant AGAN pour l'augmentation des données et CNN pour la classification de la tumeur. Cette méthode a obtenu 89,17 % résultats exacts mais 19,41% de taux de faux positifs de classification.[93],

En 2020, R. Agarwal et certains a proposé une méthode pour détecter les tumeurs cancéreuses du sein en utilisant un RCNN plus rapide sur OMI-H, Ensembles de données OMI-GE et INbreast. Les résultats obtenus en utilisant cette méthode a une sensibilité de 93 % pour OMI-H, une sensibilité de 91 % pour OMI-GE, et 99 % de sensibilité pour INbreast mais FPR et Les taux de FNR pour l'ensemble de données OMI-GE augmentent à 12 % et 20 %, pour l'ensemble de données INbreast passe à 20 % et 29 %, et le taux de FNR pour OMI-H augmente à 13 %, respectivement [94].

En 2018, L. G. Falconi et autres ont eu à proposer une méthode pour classer le cancer du sein tumeurs utilisant le réseau VGG-16 sur le CBIS-DDSM base de données. Cette méthode a atteint une précision de 82 %, mais l'inconvénient est que leur valeur MCC est de 63 %, FPR est élevé, 22 %, et le FNR est de 15 % [95].

En 2019, M. Amalmary and A. Prakash participent aux travaux en proposant une méthode pour identifier le cancer du sein dans les mammographies. Réduction de bruit, la segmentation et le regroupement sont les étapes de la proposition maquette. Le filtre gaussien est utilisé pour

éliminer le bruit de la mammographie images. Alors, flou C signifie que le regroupement est utilisé pour la segmentation de la tumeur mammaire. Le réseau de mémoire à long terme bidirectionnel (bi-LSTM) est utilisé pour diagnostiquer le cancer du sein avec des paramètres utilisant l'optimisation de l'élevage d'éléphants (EHO). Le jeu de données MIAS est utilisé dans cette recherche. Les résultats sont comparés à CNN, DCNN et Bi-LSTM, dans lequel EHOBi-LSTM a obtenu de bons résultats, mais FPR et FNR les taux doivent baisser [96].

En 2019 la proposition faite par A. M. Alqudah et ses collègues était une méthode de segmentation et la classification des tumeurs cancéreuses du sein à l'aide la technique multiseuil et PNN sur MIAS et BCDR ensembles de données. Le modèle a atteint une sensibilité de 98,30 %, un FNR de 1,7 %, et précision 97,08 %. Sa spécificité a diminué à 89,8 %, et son FPR est passé à 10,2 %. Les résultats doivent donc s'améliorer ici [97].

En 2019, N. S. Ismail and C. Sovuthy contribue aussi par une méthode de classification des tumeurs du cancer du sein en utilisant VGG et ResNet-50 sur le jeu de données IRMA. Les modèles ont atteint des précisions et les sensibilités sont de 94 % pour VGG-16, 91,7 % pour ResNet-50 % et 99 % pour VGG-16 et 94 % pour ResNet-50 %, respectivement. Sa précision pour VGG et ResNet-50 a diminué à 89 % et 88 %, respectivement. Les résultats doivent donc s'améliorer ici [98].

En 2021, R. Touahri, N. AzizI et N. E. Hammami eurent aussi à proposer une méthode de prétraitement et la classification des tumeurs cancéreuses du sein à l'aide du Algorithme LBP et modèle CNN sur le jeu de données DDSM. Le modèle a atteint une sensibilité de 96,81 %, une spécificité de 95,83 %, précision 96,32 %, FPR 4 % et FNR 3 %. Cependant, La valeur MCC tombe à 88,48 %. Donc, c'est l'écart dans leur recherches qui doivent être améliorées [99].

Tableau 2.1 : Résumé des travaux proposés

Numéro	Année	Base	Classifier	Précision	Spécificité	Sensibilité	MCC	FPR	FNR	
1	2018	CBIS-DDSM	réseau VGG-16	82 %			63%	22%	15%	
2	2019	MIAS	bi-LSTM	bonne	bonne	bonne		Baisser	Baisser	
3	2019	MIAS et BCDR	PNN	97,08 %	98,30 %	89,8 %		10,2%	1,7%	
4	2019	IRMA	VGG-16	94%		99%				
			ResNet-50	91,7%		94%				
5	2020	MIAS	CAD	92,54 %						
		données auto-collectés		95 %						
		DDSM		96,47 %						
6	2020	MIAS	SVM	90%						
			AdaBoost	57%						
			arbre de décision	54%						
			régression logistique	85%						
			K plus proche voisin	76 %						
			forestiers aléatoires	61 %						
7	2020	OMI-H	RCNN	93%					13%	
		OMI-GE		91%					12%	20%
		INbreast		99%					20%	29%
8	2021	AGAN	CNN	89,17 %					19,41%	
9	2021	DDSM	CNN	96,32	95,83 %	96,81 %	88,48%	4%	3%	

2.6 Conclusion :

Une étude plus ou moins approfondie concernant les différentes étapes d'un système DAOx a été menée dans ce chapitre. Une telle étude permet de mieux comprendre la chaîne DAOx et de faire des choix appropriés pour le traitement des masses. En effet, l'étude du prétraitement dans le domaine de l'imagerie et notamment le domaine de la mammographie sert à juger de la nécessité ou pas de cette étape dans notre étude. De plus, aborder les différentes méthodes de segmentation permet d'étudier les avantages et les inconvénients de chaque méthode et par la suite nous aider à choisir la méthode la plus appropriée pour notre application. Ces différentes étapes sont des outils, à la fois, utiles et nécessaires pour réussir le diagnostic

automatique des masses mammaires . La connaissance et l'expérience reçue de cette étude dans ce chapitre nous permet d'aborder une des parties clés de notre système dans le prochain chapitre.

Chapitre 3 : Segmentation via l'algorithme k-means

3.1 Introduction :

Dans ce chapitre, tout d'abord nous parlerons de la méthode clustering dans la segmentation des images et qui a plusieurs algorithmes d'application dont le k-means et après il serait question d'une compréhension globale du concept k-means, que fait cet algorithme en général, ensuite nous évoquerons son apport dans la segmentation des masses mammaires. Pourquoi avoir choisi cet algorithme dans notre cas. Nous donnerons tous les détails nécessaires de cet algorithme afin d'avoir une meilleure compréhension de l'utilité du K-means.

Généralités :

3.1.1 le Clustering :

Le clustering est une technique de regrouper un ensemble de points de données dans différents groupes ou clusters de telle sorte que les objets d'un groupe ont une grande similarité par rapport aux objets d'un autre groupe . Il est un apprentissage non supervisé qui aide les professionnels à découvrir des schémas cachés dans un ensemble de données. Cela aide dans une situation où il n'y a pas d'étiquette de classe dans une donnée particulière. Par le regroupement la méthode est universellement applicable, différentes méthodes sont employées pour différents objectifs de regroupement. Cependant, une compréhension commune du problème de clustering et de la méthode est essentielle pour appliquer la méthode appropriée à une tâche particulière. Dans le clustering, les pixels sont regroupés en clusters, dans lesquels les pixels d'un même cluster sont plus similaires les uns aux autres qu'à ceux de différents clusters. Parmi les types de clustering utilisés dans la segmentation d'images, nous avons le clustering K-means et le clustering flou C-means.

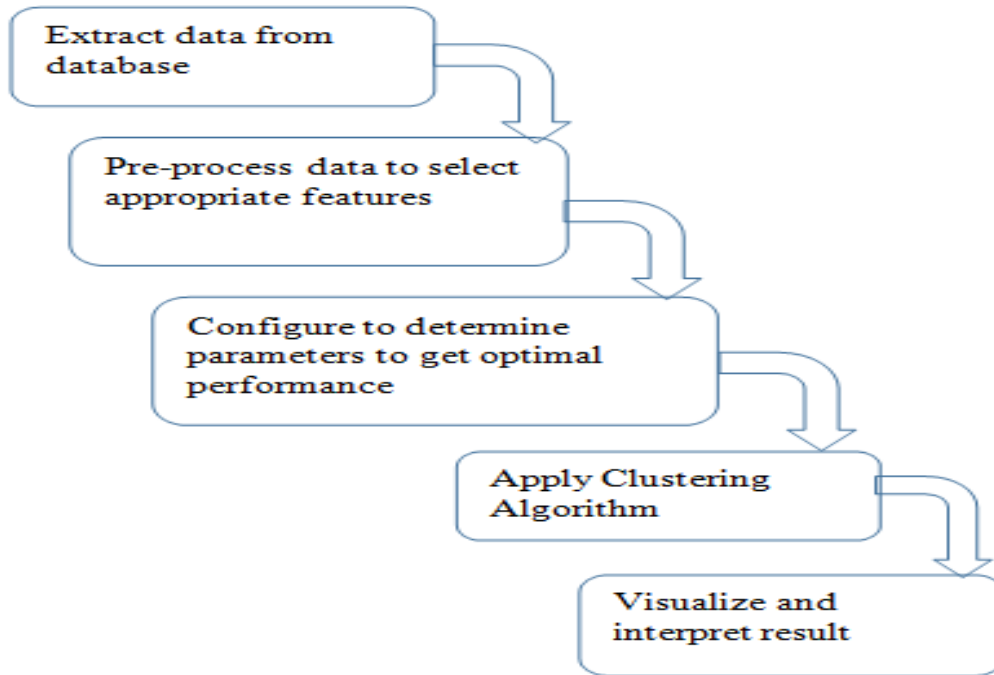


Figure 3.1 : Steps for performing clustering analysis [78]

3.1.2 L'algorithme k-means :

Description :

La méthode K-Means Clustering est un algorithme fréquemment utilisé en raison de sa facilité de mise en œuvre et il est même considéré comme le plus ancien algorithme de clustering. L'algorithme K-Means est une méthode qui peut être utilisée pour regrouper des objets jusqu'à n par similarité d'attribut dans un certain nombre de clusters k , où $k < n$. L'algorithme fonctionne en trouvant la distance la plus proche entre le centroïde et les données d'entraînement itératives.

K-means est l'un des algorithmes d'apprentissage non superviser le plus simple qui permettent de résoudre le problème de classification du fait de sa simplicité de mise en œuvre. Il est un algorithme permettant de classer ou de regrouper des objets en fonction des attributs/caractéristiques en nombre K de classe, dans laquelle les objets à l'intérieur de chaque classe sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres classes [79]. Chaque classe de la partition est définie par ses objets et son centroïde.

L'objectif est alors de minimiser la somme de l'inertie intra-classe sur l'ensemble des classes. L'algorithme procède en deux étapes : dans la première phase, on réassigne tous les objets au centroïde le plus proche, et dans la deuxième phase, on recalcule les centroïdes des classes qui ont été modifiés. Pour mesurer la proximité entre un centroïde et un objet, on

calculera une distance entre ces deux vecteurs. On pourra utiliser, par exemple, la distance euclidienne [79].

3.2 Méthode de partition du clustering en global :

L'algorithme K-means comporte deux phases différentes. Dans la première phase, les centres K (centres de gravité) sont choisis au hasard, avec une valeur fixe de K, tandis que la deuxième phase consiste à allouer chaque donnée point vers le centre le plus proche [81]. La distance euclidienne est principalement utilisée pour mesurer la distance entre les centres de cluster (centres de gravité) et chaque point de données. La distance euclidienne entre points a et b est la longueur du segment de droite reliant le point a et le point b [80].

$$\text{Distance} = \sqrt{\sum_{i=1}^n (bi - ai)^2}$$

3.2.1 Processus de clustering K-means :

k-means est un algorithme itératif qui minimise la somme des distances entre chaque individu et le centroïde. Le choix initial des centroïdes conditionne le résultat final. Admettant un nuage d'un ensemble de points, K-Means change les points de chaque cluster jusqu'à ce que la somme ne puisse plus diminuer. Le résultat est un ensemble de clusters compacts et clairement séparés, sous réserve de choisir la bonne valeur K du nombre de clusters [82].

L'algorithme K-Means est implémenté avec les étapes suivantes.

a) Initialisation Nombre de groupes

Dans cette étape, indiquez le nombre de clusters ou de segments à générer. Cette somme sera appelée k qui est un entier positif. La valeur k est généralement déterminée par une heuristique ou des études de cas .

b) Déterminer la position du premier centroïde

Sélectionnez un nombre aléatoire d'objets k dans l'ensemble de données comme centroïde initial.

c) Déterminer la distance entre les centroïdes et les autres objets

Tous les objets de données non centroïdes sont calculés en fonction de la distance par rapport à tous les centroïdes. la distance est généralement calculée par la méthode de la distance euclidienne avec l'équation suivante.

$$Distance = \sum_{j=1}^k \sum_{n \in S_j} |x_n - u_j|^2 \quad (1)$$

d) Classer l'objet dans le cluster le plus proche

Dès qu'un objet a une valeur de distance de k, entrez l'objet dans l'un des clusters les plus proches entre k .

e) Changer la position du centroïde

Pour actualiser les valeurs du centroïde, la formule suivante est utilisée.

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Une fois tous les centroïdes mis à jour, puis comparés à la valeur centroïde précédente. Si sa position ne change pas, l'algorithme se termine. Cependant, s'il y a un changement dans la position du centroïde, revenez aux processus 3 et 4. Répétez ensuite jusqu'à ce que tous les centroïdes soient stables (position inchangée). Les étapes de l'algorithme ci-dessus peuvent être illustrées avec l'organigramme de la Figure 3.2.

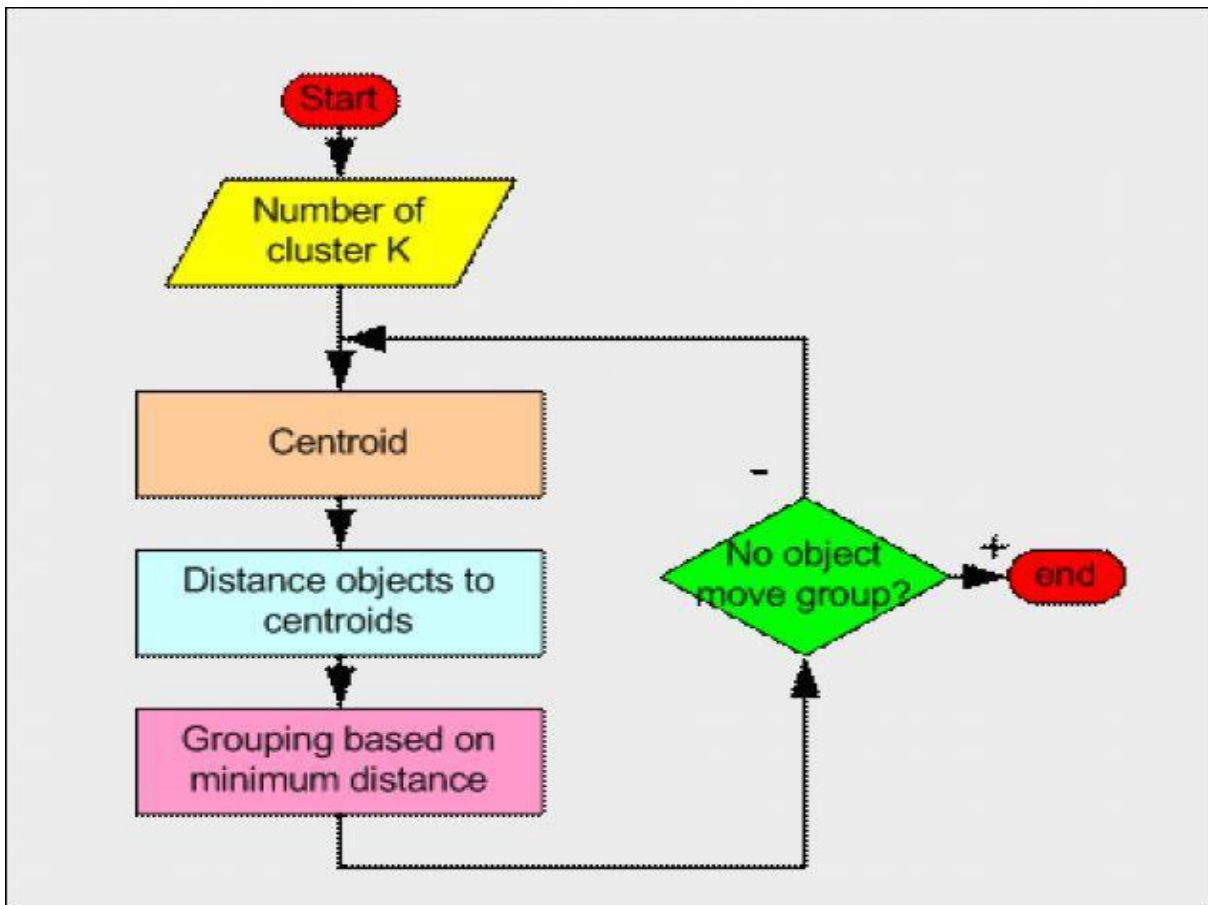


Figure.3.2 Organigramme K-means clustering [82]

3.3 Utilisation de l'algorithme k-means dans la segmentation des masses mammaires :

3.3.1 Description :

Beaucoup de recherches ont été faites sur la segmentation d'image dans différents domaines d'application, par divers techniques conventionnelles et basées sur l'apprentissage méthodes. Parmi de nombreux algorithmes de segmentation, k-means est l'un des plus simples pour cerner une région d'intérêt. Il a une complexité temporelle de $O(n)$ pour n échantillons [83]. Cependant, il est sensible aux valeurs aberrantes et les paramètres d'initialisation [84]. En réponse, il donne différents résultats de clustering avec différents clusters nombres et valeurs initiales du centroïde. Beaucoup de recherche sont penchés sur la façon d'initialiser les centres pour k-means avec l'intention de maximiser l'efficacité de l'algorithme.

Dans le clustering k-means algorithm, chaque pixel appartient à un seul cluster et centre, il s'agit donc d'un algorithme de clustering dur. Quelques travaux récents sur les méthodes de clustering de segmentation et segmentation basée sur l'apprentissage en profondeur [85]. Dans cette méthode, les pixels sont divisés en k nombres de groupes et le centre de chaque groupe est choisi parmi dont la distance est faite, les pixels proches sont formés dans un groupe et les pixels éloignés sont transformés en un autre groupe. Et par conséquent, la formation de clusters est visible. Des groupes peuvent être formés dans l'image en considérant le fond de la peau, lésions cutanées intérieures, etc. k-means est rapide par rapport à toute autre segmentation processus. Il combine n nombre de pixels pour former un nouveau cluster. Ici, la différence moyenne qui est fournie, la valeur la plus élevée donne une image de bonne qualité et l'erreur absolue normalisée est, en calculant l'erreur en image segmentée. Il doit être bas pour de meilleures performances [86].

3.3.2 Matériels et méthodes utilisés dans cette segmentation de la masse :

Cette partie est essentiellement basée sur la méthode utilisée, comment parvient-elle à segmenter l'image et identifier le tumeur ou masse. Les matériels aussi seront évoqués. A quoi consiste cette méthode dans la segmentation des masses mammaires : Tout d'abord il existe déjà plusieurs algorithmes mais ce choix a été fait car il est facile et c'est le plus utilisé dans ce cas.

L'algorithme K-means est une méthode de segmentation simple mais admirée. La vitesse d'exécution est très élevée. Mais le problème avec l'algorithme K-means est que si les centres

de cluster initiaux sont choisis de manière incorrecte, cet algorithme peut ne pas converger. Cela se produit principalement dans le cas d'une image bruyante. L'algorithme K-Means est un algorithme de clustering non supervisé qui classe les points de données d'entrée en plusieurs classes en fonction de leur distance inhérente les uns des autres. L'algorithme suppose que les caractéristiques des données forment un espace vectoriel et essaie d'y trouver un regroupement naturel.

Cet algorithme est une séparation humble et la plus utilisée algorithme de clustering basé sur de nombreuses recherches dans le monde actuel. L'algorithme des k-moyennes est un répétitif technique utilisée pour déchirer une image affectée par le cancer en k clusters. En méthode statistique et apprentissage automatique, ce regroupement est une méthode d'approche de groupe que nous pouvons diviser en n nombre de commentaires en k cluster, dans lequel chaque nombre d'observations est dans la place correcte dans le nombre de groupes avec les adjacentes moyenne. C'est le meilleur apprentissage non supervisé la plus humble technique qui est utilisée pour résoudre le rassemblement bien connu comme problème [87]. La procédure suit une méthode humble et facile pour classer un ensemble de données donné par un nombre sûr de groupes. De nombreux algorithmes sont utilisés pour le jeu de données choisi. Pour diviser n nombre de motifs $\{x_1, \dots, X_n\}$ dans un espace d-dimensionnel en k- groupes (supposons k clusters) en utilisant cet algorithme de clustering. L'ensemble de résultats sera l'ensemble des k centres, dans lequel chacun des clusters est positionné au centre de gravité du jeu de données divisé [87]. Ceci est décrit dans les étapes suivantes :

Étape 1 : Le nombre de valeurs de cluster pour k doit être saisi.

Étape 2 : Après cela, choisissez les centres du groupe k au hasard.

Étape 3 : la moyenne ou le centre du cluster est calculé.

Étape 4 : La distance entre chaque pixel à chaque centre du cluster est ensuite calculée.

Étape 5 : Si la distance est proche du centre, alors nous devrions passer à ce groupe.

Étape 6 : Sinon, passez au groupe suivant.

Étape 7 : réévaluer le point médian du cluster.

Étape 8 : Récapitulez le même processus jusqu'au milieu ne bouge pas.

De nombreux algorithmes de clustering montrent leur efficacité dans domaine différent. Cette technique est la technique fréquemment utilisée dans tous les domaines. En particulier, dans le

domaine médical, l'efficacité de k-means prouvée dans de nombreuses systèmes par ses performances.

3.4 Les méthodes de clustering les plus utilisées :

Les algorithmes de clustering sont de plusieurs types. L'aperçu suivant ne répertorie que les exemples les plus importants d'algorithmes de clustering, car il existe plus de 100 algorithmes de clustering publiés. Tous ne fournissent pas de modèles pour leurs clusters et ne peuvent donc être facilement catégorisés.

3.4.1 Fuzzy c-means algorithm :

L'algorithme (FCM) est une technique de clustering flou en vue de la minimisation d'un paradigme quadratique où les clusters sont représentés via leurs centres séparés. L'algorithme FCM a été proposé par Dunn en 1973 et amélioré par Bezdek [88] en 1981. Pour un groupe de motifs d'information C $X=\{x_1,x_2,\dots,x_N\}$ l'algorithme permet de segmenter l'espace d'information, en chiffrant les centres de classes et la matrice d'appartenance, de plus, en limitant une fonction objectif J concernant ces axes et degrés d'appartenance comme suit :

$$J=\sum_{i=1}^N \sum_{j=1}^C \mu_{ik}^m \|x_i - c_j\|^2 \quad (1)$$

J est la fonction objectif. Après une itération de l'algorithme l'estimation de J est plus petite qu'auparavant. Cela implique que l'algorithme rejoint ou se rapproche d'une division décente de pixels en groupes. C 'est le nombre de clusters utilisés dans le cadre de l'algorithme, et doit être choisi avant l'exécution, N est la quantité de pixels dans l'image, est la table d'appartenance - une table de $N \times C$ sections qui contient les estimations d'appartenance de chaque point de référence et chaque cluster [89]. μ_{ik}^m est le degré d'appartenance de x_i au cluster j , m est un facteur de flou (une estimation supérieure à 1), x_i est le i ème pixel de N , c_j est le j ème cluster de C et la différence entre $(x_i - c_j)$ est la Distance euclidienne entre x_i et x_j [90].

3.4.2 Méthodes basées sur la distribution :

Il est question d'un modèle de clustering dans lequel des données sont ajoutées sur a probabilité selon laquelle elles peuvent appartenir à la même distribution. Le regroupement effectué est soit normal ou Gaussien. La distribution Gaussienne est plus importante lorsqu'il y a un nombre fixe de distributions et que toutes les données à venir y sont intégrées de sorte que la distribution des données puisse être maximisée. Il en résulte un regroupement qui est illustré dans la figure suivante :

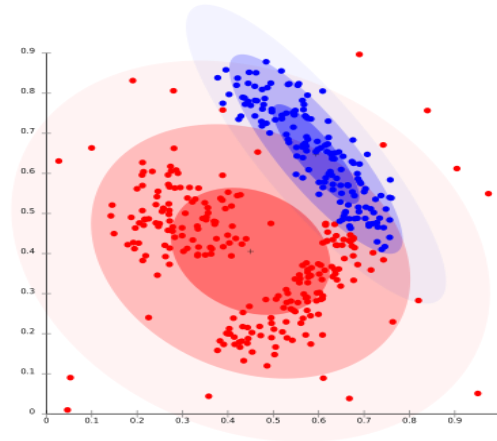


Figure 3.3 : L' algorithme de maximisation des attentes [91]

Ce modèle fonctionne bien sur des données synthétiques et des clusters de tailles diverses. Mais ce modèle peut avoir des problèmes si les contraintes ne sont pas utilisées pour limiter la complexité du modèle. De plus, le clustering basé sur la distribution produit des clusters qui supposent des modèles mathématiques définis de manière concise sous-jacents aux données, une hypothèse assez forte pour certaines distributions de données. Pour Ex- L' algorithme de maximisation des attentes qui utilise des distributions normales multivariées est l'un des exemples populaires de cet algorithme.

3.4.3 Méthodes basées sur le centroïde :

il s'agit essentiellement de l'un des algorithmes de clustering itératifs dans lesquels les clusters sont formés par la proximité des points de données avec le centroïde des clusters. Ici, le centre du cluster, c'est-à-dire le centroïde , est formé de telle sorte que la distance des points de données soit minimale avec le centre. Ce problème est fondamentalement l'un des problèmes NP-Hard et donc les solutions sont généralement approximées sur un certain nombre d'essais. Pour Ex- K – signifie que l'algorithme est l'un des exemples populaires de cet algorithme [91].

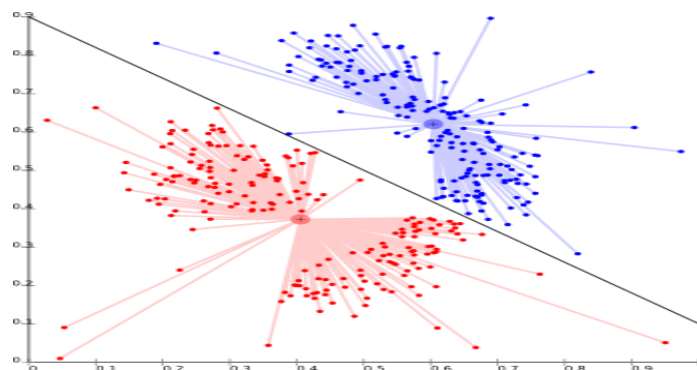


Figure 3.4 : k-means [91]

Le plus gros problème avec cet algorithme est que nous devons spécifier K à l'avance. Il a également des problèmes pour regrouper les distributions basées sur la densité [91].

3.4.4 Méthodes basées sur la connectivité :

L'idée centrale du modèle basé sur la connectivité est similaire au modèle basé sur le centroïde qui définit essentiellement les clusters sur la base de la proximité des points de données. Ici, nous travaillons sur une notion selon laquelle les points de données qui sont plus proches ont un comportement similaire par rapport aux points de données qui sont plus éloignés. Il ne s'agit pas d'un partitionnement unique de l'ensemble de données, mais plutôt d'une hiérarchie étendue de clusters qui fusionnent les uns avec les autres à certaines distances. Ici, le choix de la fonction de distance est subjectif. Ces modèles sont très faciles à interpréter mais manquent d'évolutivité [91].

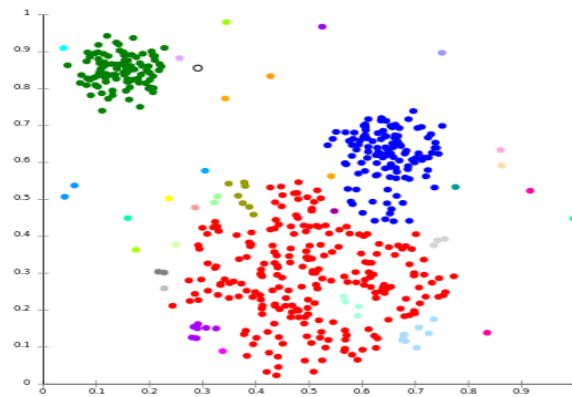


Figure 3.5 : algorithme hiérarchique et ses variantes [91].

3.4.5 Modèles de densité :

Dans ce modèle de clustering, il y aura une recherche d'espace de données pour les zones de densité variée de points de données dans l'espace de données. Il isole diverses régions de densité en fonction des différentes densités présentes dans l'espace de données [91].

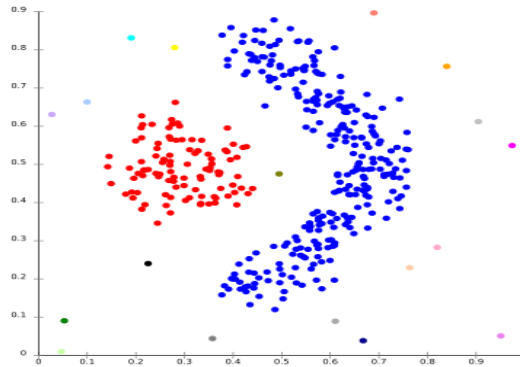


Figure 3.6 : DBSCAN et OPTICS [91].

3.4.6 Agrégation de sous-espace :

Le clustering de sous-espace est un problème d'apprentissage non supervisé qui vise à regrouper des points de données en plusieurs clusters de sorte que les points de données d'un seul cluster se situent approximativement sur un sous-espace linéaire de faible dimension. Le regroupement de sous-espaces est une extension de la sélection de caractéristiques, tout comme le regroupement de sous-espaces de sélection de caractéristiques nécessite une méthode de recherche et des critères d'évaluation, mais en plus, le regroupement de sous-espaces limite la portée des critères d'évaluation. L'algorithme de regroupement des sous-espaces localise la recherche des dimensions pertinentes et leur permet de trouver le cluster qui existe dans plusieurs sous-espaces qui se chevauchent. Le clustering de sous-espace était à l'origine destiné à résoudre des problèmes de vision par ordinateur très spécifiques ayant une union de structure de sous-espace dans les données, mais il attire de plus en plus l'attention de la communauté des statistiques et de l'apprentissage automatique. Les gens utilisent cet outil dans les réseaux sociaux, les recommandations de films et les ensembles de données biologiques. Le regroupement de sous-espaces soulève le problème de la confidentialité des données, car de nombreuses applications de ce type impliquent le traitement d'informations sensibles. Les points de données sont supposés être incohérents car ils ne protègent que la confidentialité différentielle de toute caractéristique d'un utilisateur plutôt que l'ensemble du profil utilisateur de la base de données [91].

Il existe deux branches de regroupement de sous-espaces en fonction de leur stratégie de recherche :

- Les algorithmes descendants trouvent un clustering initial dans l'ensemble complet des dimensions et évaluent le sous-espace de chaque cluster.
- L'approche ascendante trouve une région dense dans un espace de faible dimension puis se combine pour former des clusters [91].

3.6 Conclusion :

Il y a beaucoup d'autres méthodes de segmentation plus précisément de clustering mais juste il est simple et beaucoup utilisé d'où ce choix. Grace à la segmentation qui est une partie essentielle du système, nous pourrions plus facilement aborder la classification et la simplifier. Les images mammographiques malignes et bénignes sont segmentées par l'algorithme k-Means pour faciliter l'identification de la zone tumorale du cancer du sein afin de procéder au diagnostic.

Dans cette partie nous avons eu à faire un détail complet de la méthode de segmentation qui est le k-means pourquoi l'avons choisi parmi tant d'autres cela ne montre pas qu'il est le meilleur, mais rapide par rapport aux autres méthodes de segmentation ce qui est sa plus grande particularité. L'approfondissement de son étude facilite son implémentation dans la conception du système. Ce qui nous amène dans notre prochain chapitre dans lequel seront détaillées la conception et l'implémentation de notre système.

Chapitre 4 : Conception et Implémentation du système

4.1 Introduction :

La première partie concernera les différentes étapes de conception de notre système d'aide au diagnostic médical. Chaque partie du système sera détaillée et sa participation c'est-à-dire ce qu'elle fait et la seconde partie montrera le déroulement de notre système, une démonstration. Ceci comportera la sélection de l'image mammographique jusqu'à la segmentation complète ainsi que l'extraction des caractéristiques du premier ordre et enfin la classification. Nous donnerons des visuels pour la compréhension.

4.2 Conception du Système :

4.2.1 Environnement et outils de travail :

- **Ordinateur**

L'intégralité des expérimentations décrites dans notre travail a été réalisée sur un ordinateur (Asus) est essentiellement configuré avec Intel Processeur (R) Intel(R) Core (TM) i3-4005U CPU @ 1.70GHz 1.70 GHz et 4 Go de mémoire en cours d'exécution.

- **ANACONDA**

Nous avons décidé de travailler avec le logiciel ANACONDA contenant plusieurs fonctionnalités dont SPYDER 3.9.7 qui nous facilite la programmation en langage PYTHON 3.9.7.

- **WEKA**

WEKA (waikato Environment for knowledge analysis) est une collection d'algorithmes d'apprentissage automatique pour les tâches d'exploration de données. Les algorithmes peuvent être appliqués directement à un ensemble de données ou appelés à partir de votre propre code Java. WEKA contient des outils pour le prétraitement des données, la classification, la régression, le clustering, les règles d'association et la visualisation. Il est également bien adapté au développement de nouveaux schémas d'apprentissage automatique.

- **Eclipse**

Eclipse est un IDE (*Integrated Development Environment*), (EDI environnement de développement intégré en français), c'est-à-dire un logiciel qui simplifie la programmation en

proposant un certain nombre de raccourcis et d'aide à la programmation. Il est développé par IBM, est gratuit et disponible pour la plupart des systèmes d'exploitation [100].

- **Langages utilisés :**

Ce système est fait pratiquement du langage **python** accompagné du **java**. Ce langage est grâce à sa facilité et sa maniabilité. La principale raison est qu'elle contient déjà des fonctionnalités qui nous facilite l'implémentation et permet de repérer vite les erreurs et les corriger. Il possède des méthodes déjà disponibles, il suffit juste de les importer ce qui donne un bon avantage.

4.2.2 Images Mammographiques (MIAS) :

The Mammographic Image Analysis Society (MIAS) est une organisation regroupant des groupes de recherche britanniques qui s'intéressent à la compréhension des mammographies. Elle a créé une base de données de mammographies numériques. Les films prélevés dans le cadre du programme national britannique de dépistage du cancer du sein ont été numérisés jusqu'à un bord de pixel de 50 microns. La base de données contient 322 films numérisés. Elle comprend également les marques de "vérité" du radiologue sur l'emplacement de toute anomalie éventuelle. La base de données a été réduite à un bord de pixel de 200 microns et clippée de sorte que toutes les images soient 1024x1024. Les images mammographiques sont disponibles via le Pilot European Image Processing Archive (PEIPA) à l'Université d'Essex [101].

Nous avons utilisé 46 images tirées de la base de données MIAS, où 24 sont de type bénin et 22 de type malin.

4.2.3 Métriques d'évaluation :

Afin d'évaluer notre travail, nous nous sommes basés sur le calcul de trois métriques d'évaluation : la précision, la sensibilité et la spécificité.

- **La précision** d'un test est sa capacité de différencier correctement les cas de tumeurs malignes et bénignes.

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

- **La sensibilité** d'un test est sa capacité de déterminer correctement les cas de malins.

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

- **La spécificité** d'un test est sa capacité de déterminer correctement les cas bénins.

La spécificité = $VN/(VN+FP)$

- **Matrice de confusion** appelée aussi tableau de contingence. Elle mettra non seulement en valeur les prédictions correctes et incorrectes mais nous donnera surtout un indice sur le type d'erreurs commises. Pour calculer une matrice de confusion, on a besoin d'un ensemble de données de test et un autre de validation qui contient les valeurs des résultats obtenus.

Chaque colonne du tableau contient une classe prédite par l'algorithme et les lignes des classes réelles [102].

		Classe réelle	
		-	+
Classe prédite	-	True Negatives <i>(vrais négatifs)</i>	False Negatives <i>(faux négatifs)</i>
	+	False Positives <i>(faux positifs)</i>	True Positives <i>(vrais positifs)</i>

Figure 4.1 : Matrice de confusion

Où :

Vrai positif (VP) = Les cas correctement classés comme malignes.

Faux positif (FP) = Les cas incorrectement classés comme malignes.

Vrai négatif (VN) = Les cas correctement classifiés comme bénignes.

Faux négatif (FN) = Les cas incorrectement classés comme bénignes.

4.2.1 L'Organigramme du système :

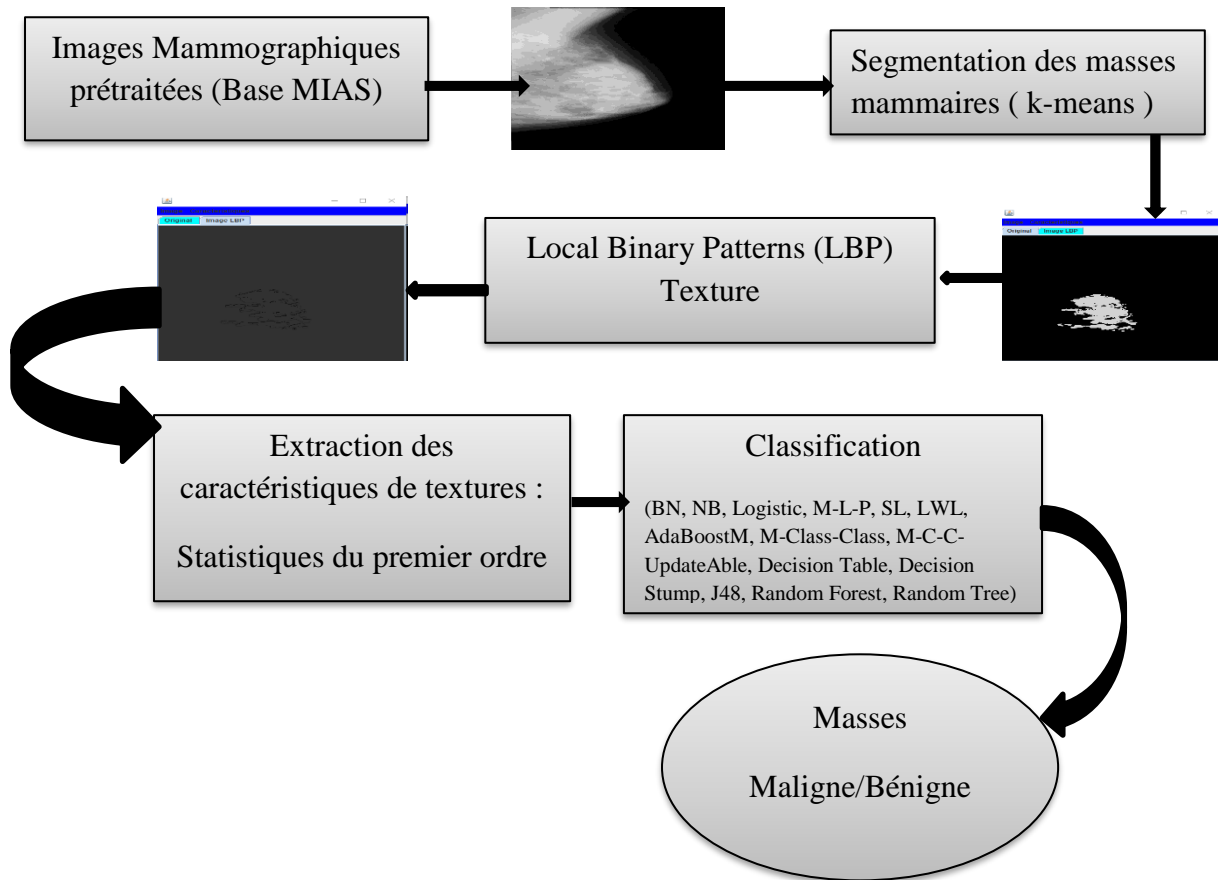


Figure 4.2 : Organigramme de notre Système CADx (Computer Aided Diagnosis)

Le système proposé, ici, a comme principale fonction de répondre aux besoins du diagnostic du cancer du sein nécessaire dans la pratique médicale courante. Le médecin est de ce fait le principal utilisateur de cette application.

Il offre au médecin la garanti d'avoir les meilleurs résultats en un temps record, cela permet l'efficacité, la rapidité dans la recherche du tumeur et la prise en charge du patient ou plusieurs. Il détache le médecin de certaines taches lui ralentissant dans son travail car beaucoup de recherches et pertinences des résultats.

4.2.5 La segmentation par la méthode k-means :

Après avoir choisi l'image à segmenter, nous appliquons maintenant notre méthode de segmentation qui n'est autre que le clustering k-means qui permet de segmenter l'image en plusieurs segments jusqu'à trouver le tumeur. Nous aurons des clusters avant le cluster final.

Comment fonctionne ? nous allons tout d'abord donner le nombre de cluster k et le nombre de déroulement runs. L'algorithme se déroule et segmente l'image en plusieurs tout en tenant compte de l'intensité du pixel la plus élevée, pour cela améliorons l'intensité de notre image cela permet une bonne visualisation du tumeur ce qui donne une idée sur la partie à segmenter . après avoir trouvé le tumeur, nous prenons le dernier cluster qui représente la finalité de notre segmentation.

Vue qu'on a juste besoin de la partie tumeur et non la totalité de l'image segmentée, nous allons l'extraire et masquer le reste de l'image donc on n'a pas besoin. Ceci met fin à notre segmentation .place à l'extraction des caractéristiques.

4.2.5.1 Expériences et résultats :

Utilisons notre de bases d'images pour des expériences. Les résultats expérimentaux sont illustrés à la figure 4.3

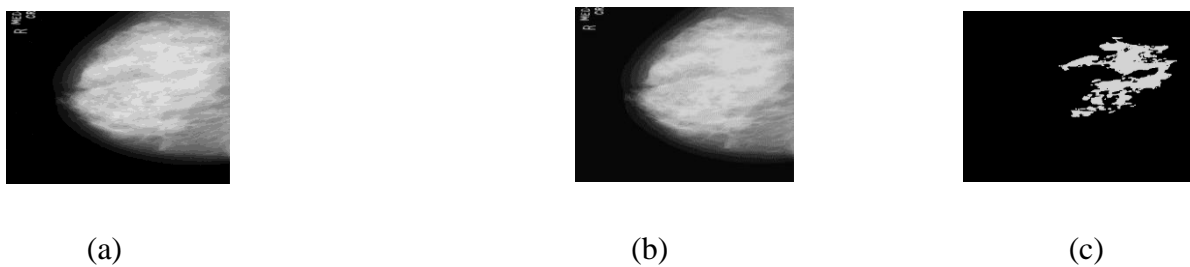
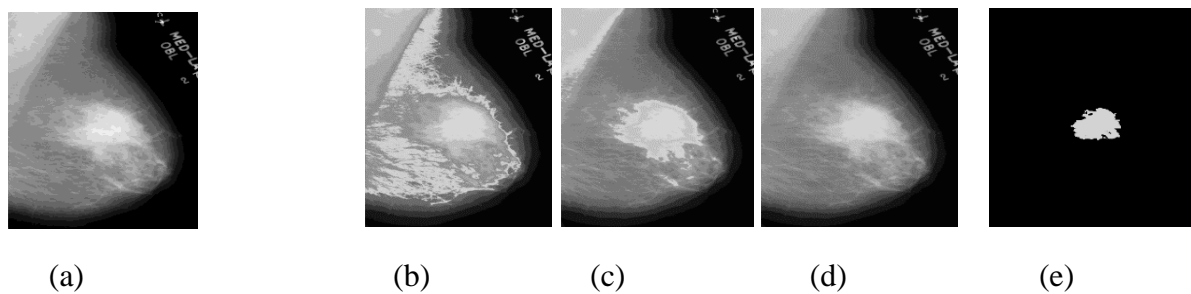


Figure 4.3 : résultats de la segmentation d'images par k-means, (a) image originale 236.png du MIAS, (b) image segmentée par k-means, (c) extraction du tumeur(masse mammaire) .

4.2.5.1.1 k valeurs :

Afin de vérifier davantage l'efficacité de la méthode proposée, plusieurs autres groupes d'images sont sélectionnés et testés avec différentes valeurs de K . Les résultats expérimentaux sont présentés dans la figure 4.4 :



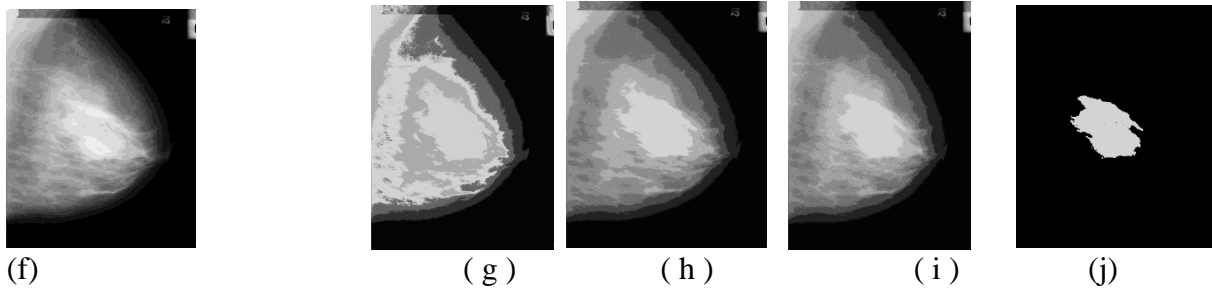


Figure 4.4 : segmentation avec k valeurs ; (a), (f) originales (81B.png et 315B.png) ; (b),(c),(d) segmentation k=10,11,12 respectivement ; (f),(g),(h) segmentation k=6,7,8 respectivement ; (e),(j) résultats k-means .

4.2.5.2 Explication des résultats :

Sachons que le clustering k-means regroupe en fonction de l'intensité du pixel. Après segmentation, les résultats se présentent dans la figure 3.4. Il est facile d'identifier que certains des groupes ont moins de pixels en fonction de l'intensité et les normes maximales de pixels peuvent être obtenues dans certains autres groupes. Il est facile d'identifier que le cluster (d) et (h) ont quelques différences dans l'intensité et une précision dans la segmentation pour indiquer la partie la plus affectée de l'image du cancer. Selon les suggestions des médecins praticiens, donc notre expérience nous pourrions conclure que les régions touchées se trouvent dans les groupes (d) et (h) [100]. Après une bonne précision dans la segmentation, donc on s'arrête et on essaie de récupérer juste le tumeur. (e) et (i) représentent l'extraction du tumeur sans le reste de l'image mammographique. Ce qui conclut notre segmentation à base du k-means .

4.2.6 Extraction des caractéristiques :

Récupérons notre image segmentée par le biais d'un sauvegarde automatique après segmentation, appliquons l'algorithme LBP (Local Binary Patterns) qui est couramment utilisé pour agir sur la texture d'une image, donc cela nous permettra d'avoir une bonne vision des différents textures que contiennent notre image segmentée qui est la partie tumeur, à la suite de cela récupérons cette image dont la texture est visible grâce à LBP et appliquer les descripteurs statistiques du premier ordre qui contient plusieurs mesures. Cela permet encore plus une bonne analyse des textures.

4.2.7 Classification :

Cette partie est la finalité du système, elle permet de récupérer le résultat de tous ces traitements sur l'image choisie et de la classifier, nous nous sommes basés sur plusieurs techniques de classification qui donnent une précision sur le tumeur indiquant si elle est maligne

ou bénigne dans notre cas. Ce qui permettra un éventuel traitement du patient se basant sur ce résultat.

4.3 Implémentation du système :

Partie 1 : Segmentation

4.3.1 Interface :

Elle est composée de deux fenêtres, la première étant comme une fenêtre d'accueil et le bouton START nous mène à la deuxième qui contient deux Canvas un pour l'image originale et l'autre pour le résultat des différents traitements de l'image.

- Première fenêtre :



Figure 4.5 : Première fenêtre

- Deuxième fenêtre :

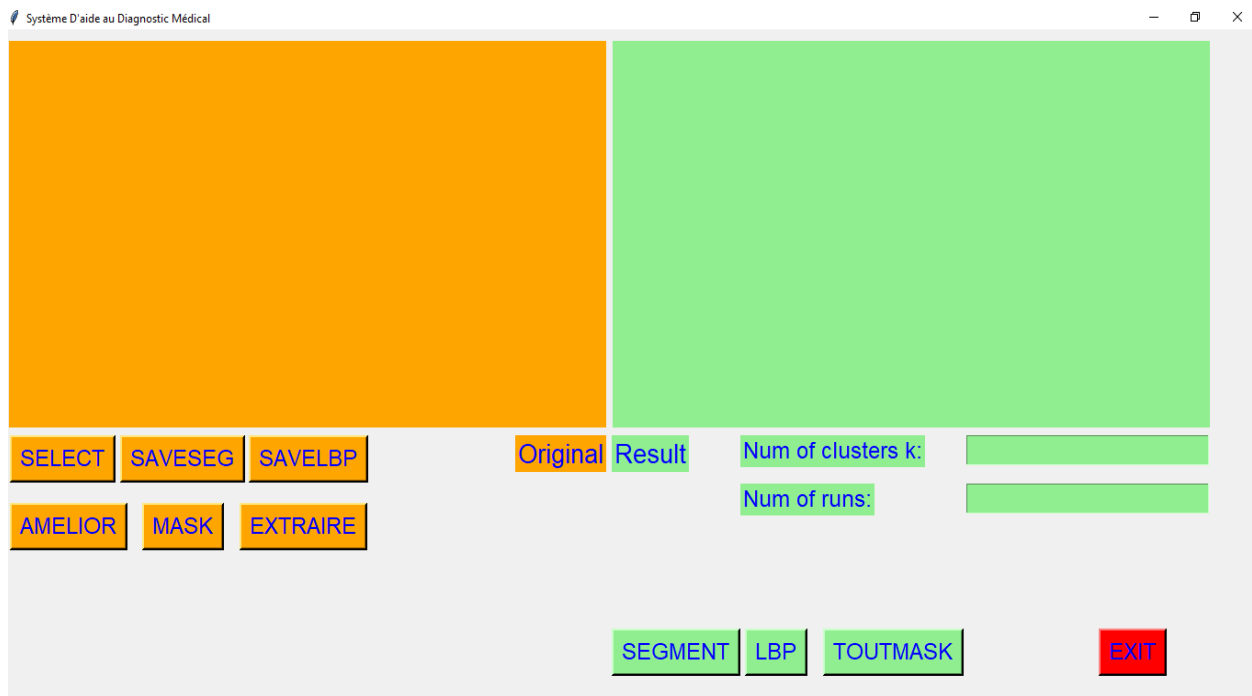


Figure 4.6 : fenêtre d'accueil

4.3.2 Sélection de l'image :

Le bouton **SELECT** nous permet cela

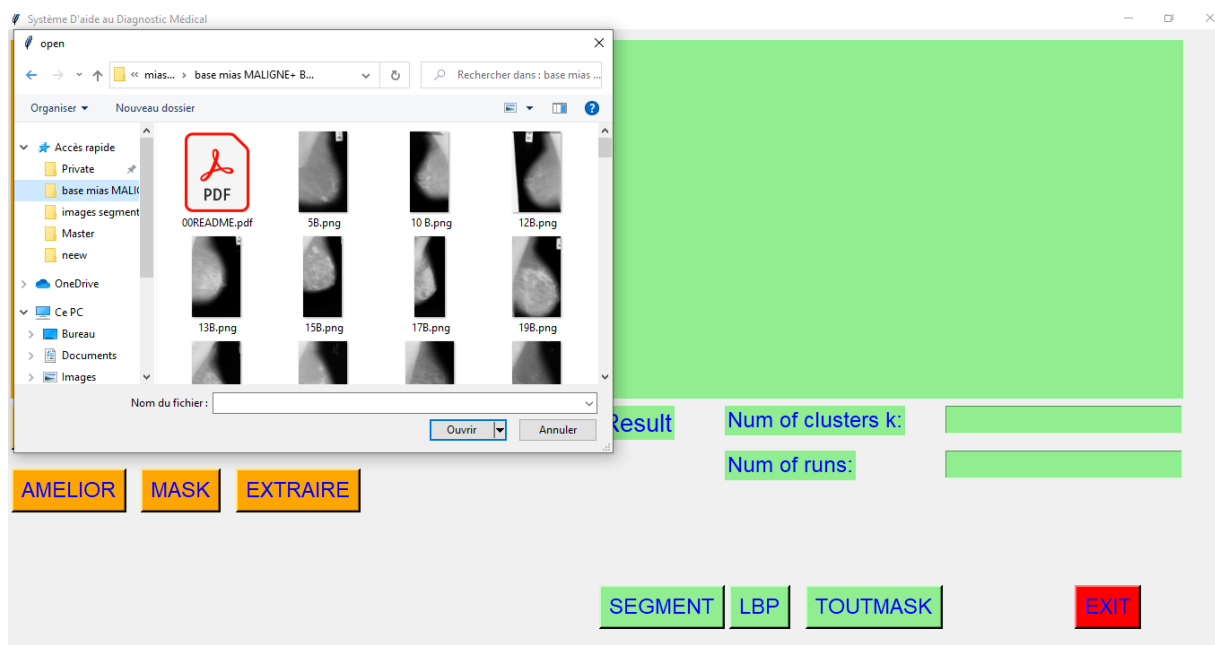


Figure 4.7 : choix de l'image

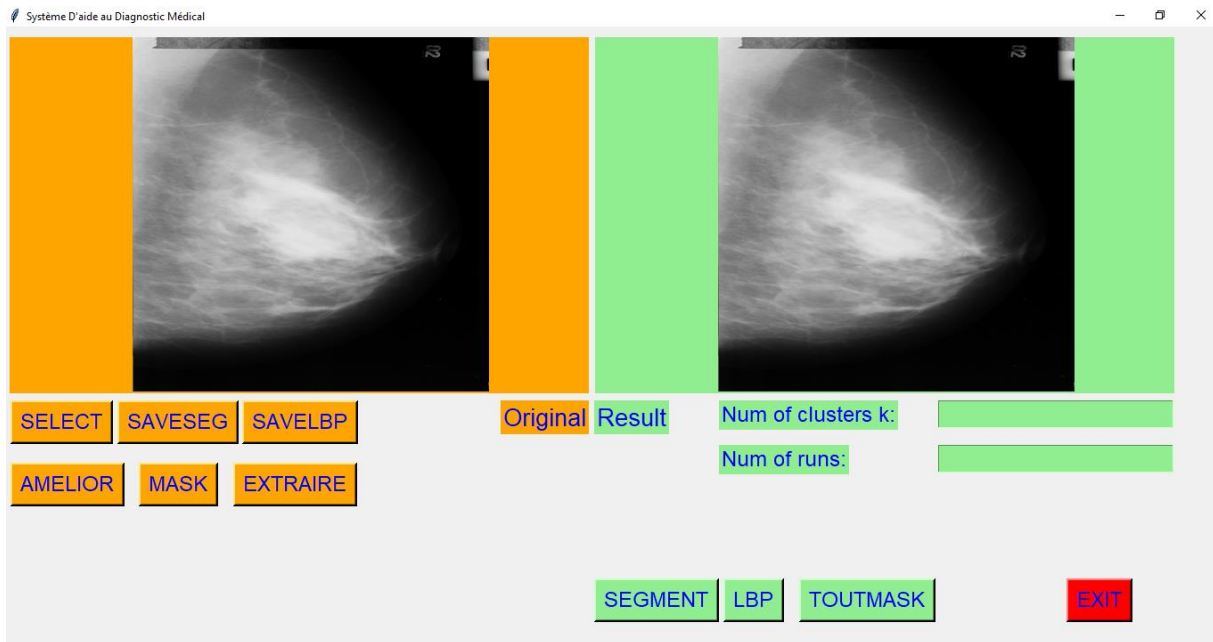


Figure 4.8 : image choisie

4.3.3 Segmentation image :

Tout d'abord nous devons introduire le nombre de k et le nombre de runs ensuite cliquer sur le bouton SEGMENT pour faire la segmentation ainsi de suite pour trouver le tumeur.

Que fait le k-means, il segmente l'image en plusieurs partie et densifie l'intensité en une partie pour l'extraction

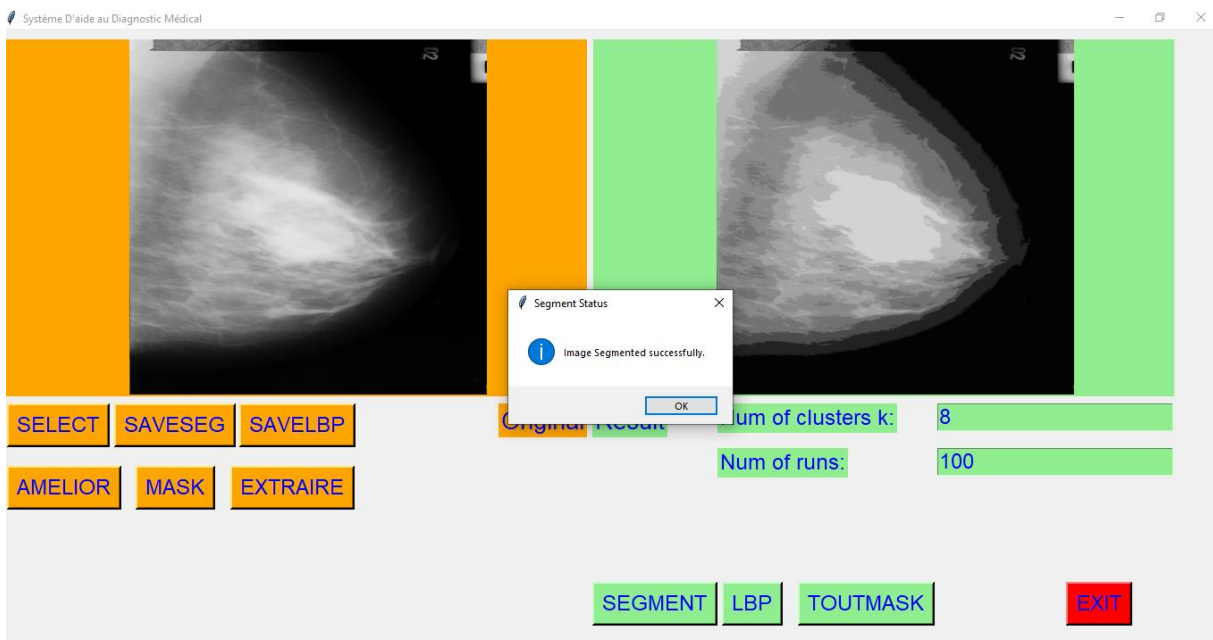


Figure 4.9 : Segmentation 315B.png

Faisons une extraction du tumeur, nous allons charger l'image qui vient d'être segmentée qui est enregistrée automatique et appuyons sur EXTRAIRE. Cette extraction prend les parties dont l'intensité du pixel est élevée après segmentation et ensuite met les autres en noir

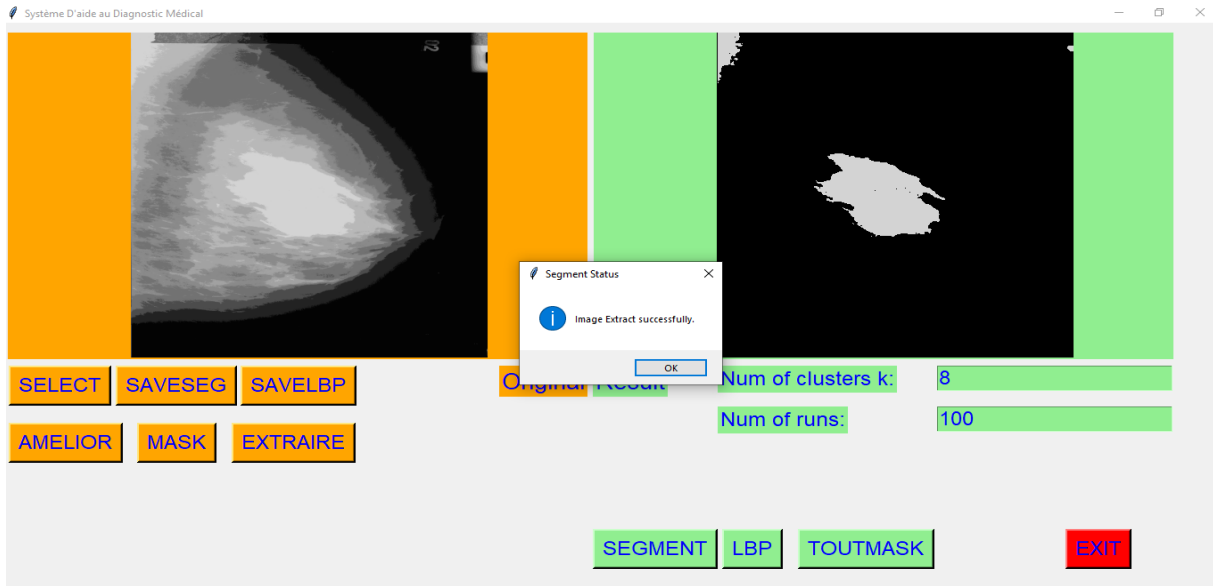


Figure 4.10 : tumeur extraite

Ensuite le traitement continue, faisons le contour avec MASK à l'image extraite et terminons par le TOUTMASK sur l'image obtenue après contours qui prend le tumeur seul en compte car l'autre partie visible sur l'image n'est pas tumeur mais juste une glande du sein donc on le masque en noir il restera juste le gris que l'on recherche



Figure 4.11: extraction complète du tumeur

Partie 2 : LBP, Extraction de caractéristiques et classification

Cette partie comporte deux phases : l'apprentissage et la classification.

▪ **La phase d'apprentissage**

- Dans cette partie, nous nous basons sur les 46 images, et nous suivons les étapes suivantes:
- Création de la base d'apprentissage «CAD-LBP.arff » (voir figures 4.12 et 4.13).
- Chargement de l'image qui contient la région d'intérêt (voir figure 4.14).
- Application de la méthode LBP (voir figure 4.15).
- Calcul des caractéristiques statistiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité et la régularité) et construction des vecteurs de caractéristiques (voir figures 4.16).

Pour chaque image : une fois le calcul de toutes les caractéristiques est obtenu, le vecteur qui contient les caractéristiques du premier ordre (la moyenne, la variance, l'asymétrie, l'aplatissement, l'entropie, l'uniformité, la régularité) est construit.

- Remplissage de la base d'apprentissage (voir figures 4.17).

Chaque vecteur de caractéristiques construit sera rajouté dans la base d'apprentissage pour l'envoyer à la dernière étape « la classification ».



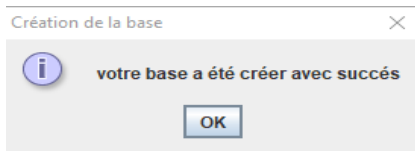


Figure 4.12 : représente la création de la base

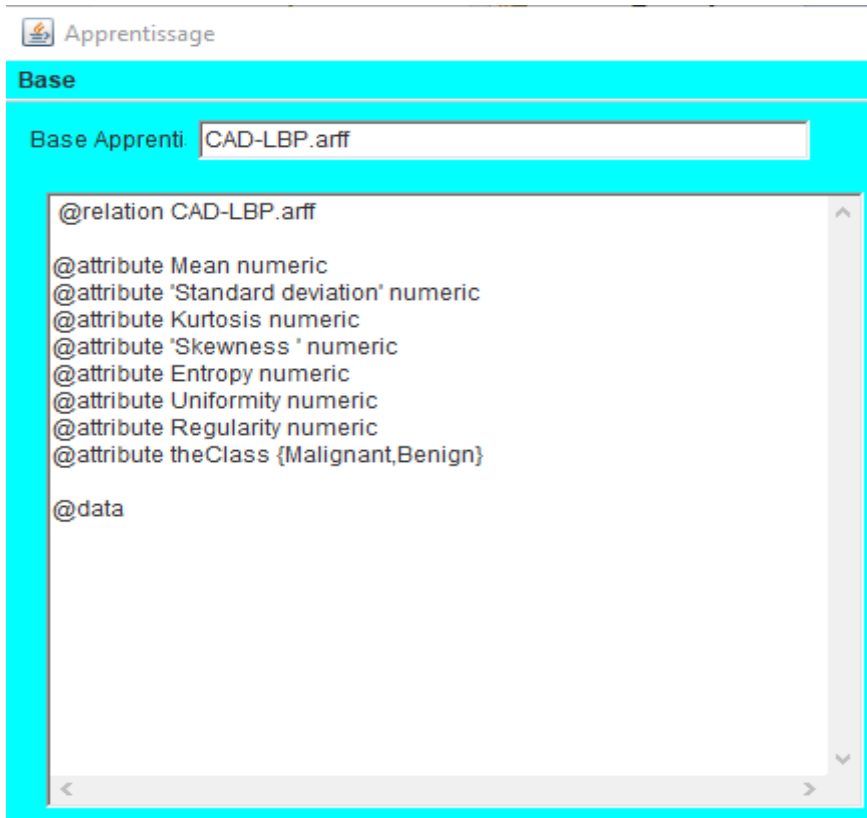


Figure 4.13 : base CAD-LBP

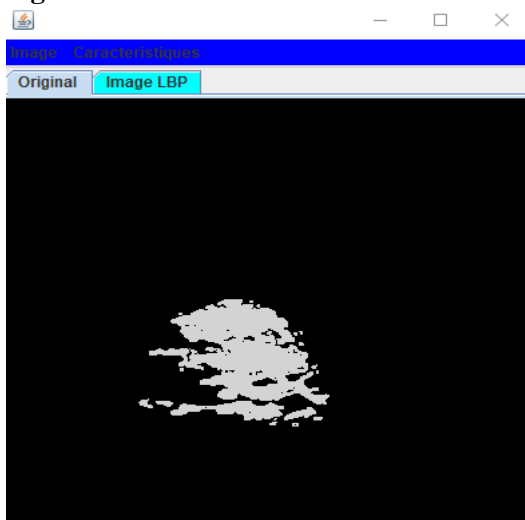


Figure 4.14 : le chargement de l'image qui contient la région d'intérêt

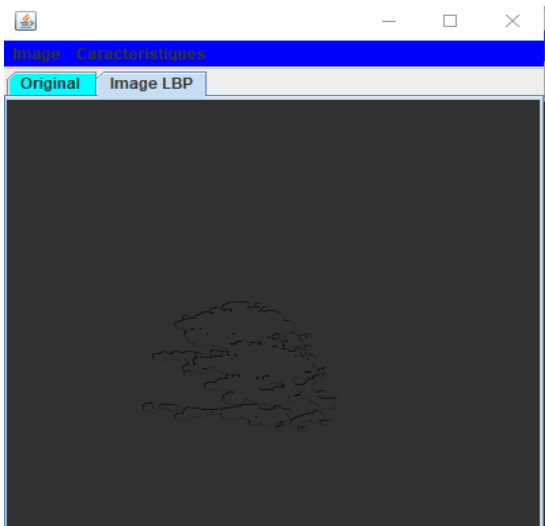


Figure 4.15 : Application de LBP sur la région d'intérêt

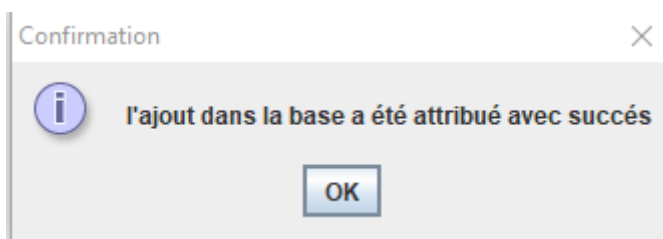


Figure 4.16 : l'ajout d'un vecteur de caractéristiques dans la base

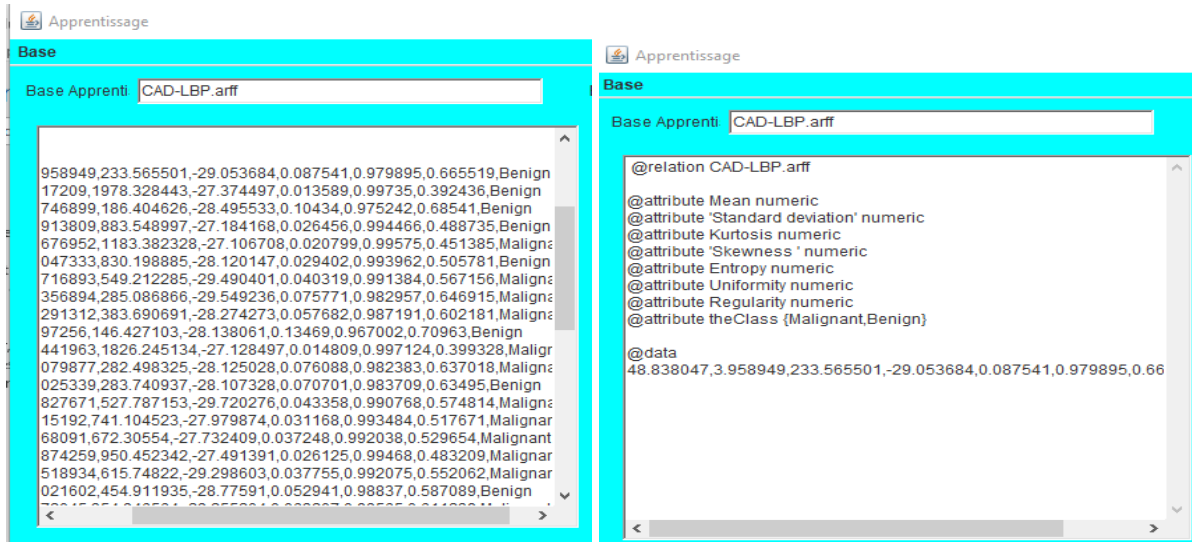


Figure 4.17 : Base d'apprentissage complète avec tous les vecteurs de caractéristiques

▪ **La phase de classification**

- Pour visualiser clairement les performances de notre travail, dans l'étape de « classification », nous avons testés plusieurs techniques de classifications.
- Nous avons utilisé quatorze classifieurs disponibles via l'outil Weka (Figure 4.18) qui possède de plusieurs classifieurs qui nous permettra de classer les tumeurs pour différencier les malignes d'avec les bénignes. Voilà les classifieurs utilisés : BayesNet, NaiveBayes, Logistic, Multi-Layer-Perception, Simpe Logistic, LWL, AdaBoostM, Multi-Class-Classifier, Multi-Class-Classifier-UpdateAble, Decision Table, Decision Stump, J48, Random Forest, Random Tree; où la valeur de la validation croisée k-fold =10.



Figure 4.18 : Interface principale de WEKA

Les figures suivantes représentent les résultats de la classification en termes de précision, sensibilité et spécificité :

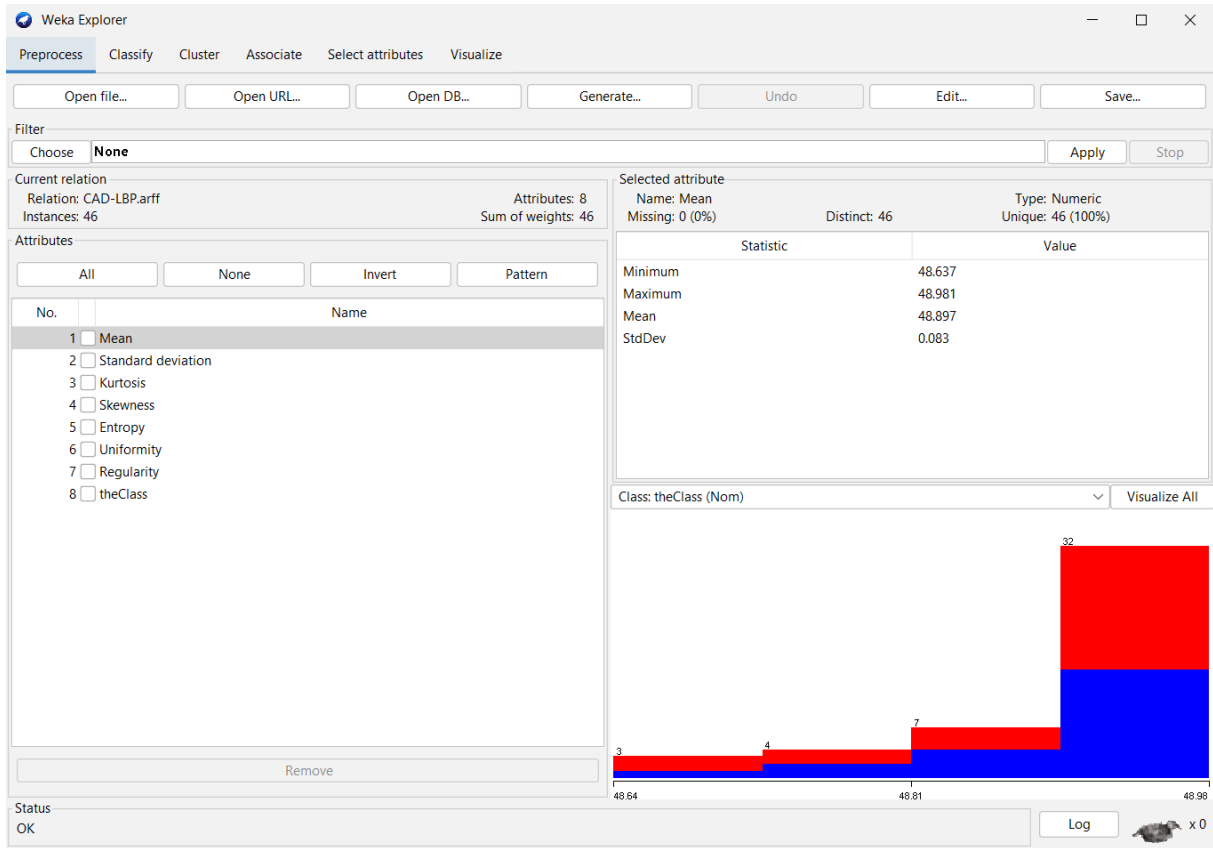


Figure 4.19 : Chargement de la base d'apprentissage

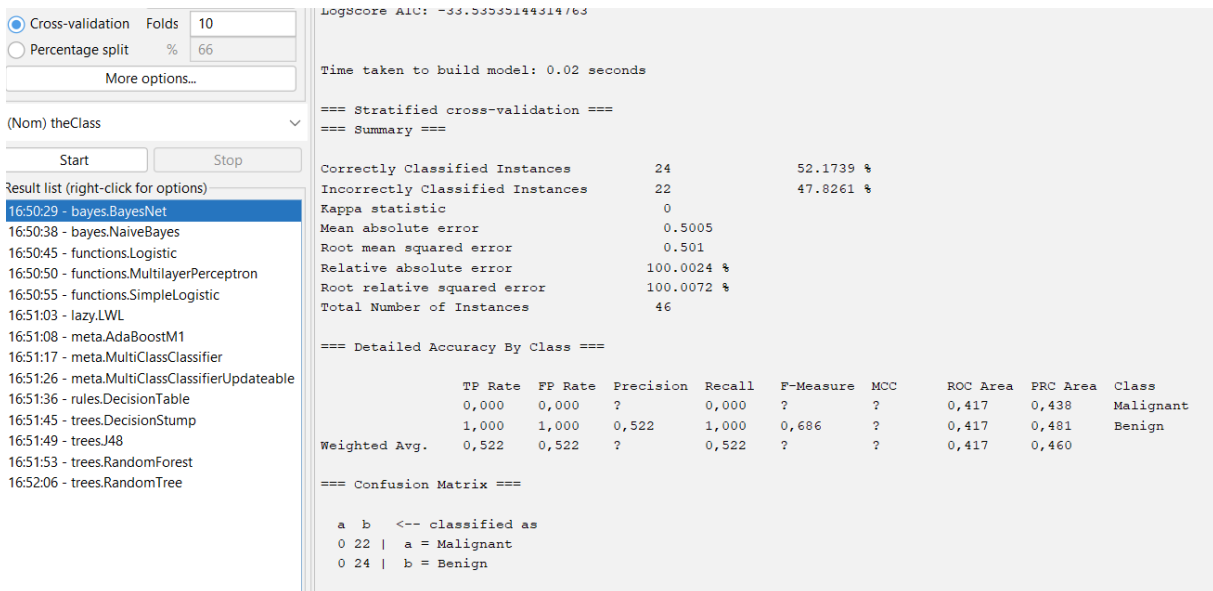


Figure 4.20 : Résultats de classification obtenus par le classifieur Bayes Network

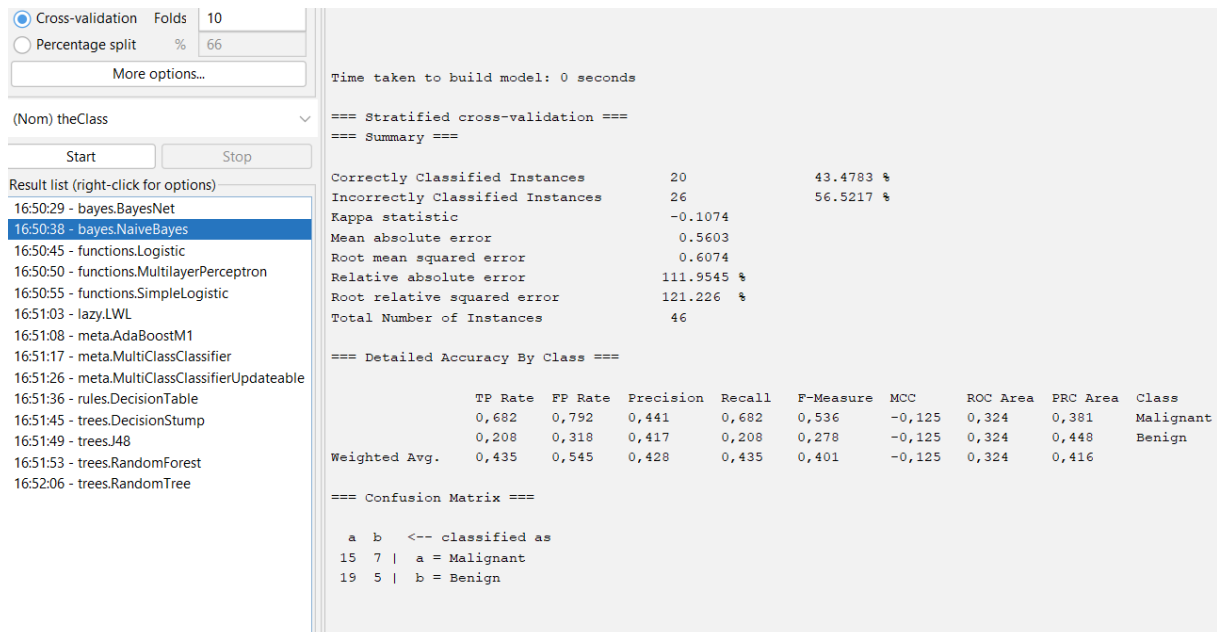


Figure 4.21 : Résultats de classification obtenus par le classifieur Naive Bayes

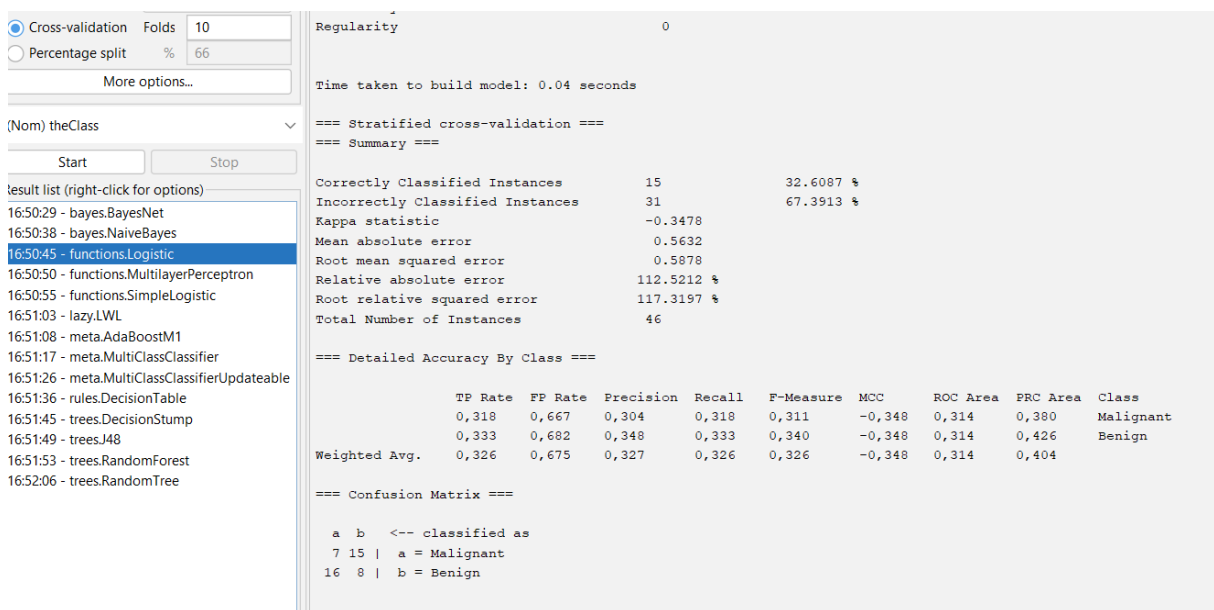


Figure 4.22 : Résultats de classification obtenus par le classifieur Logistic

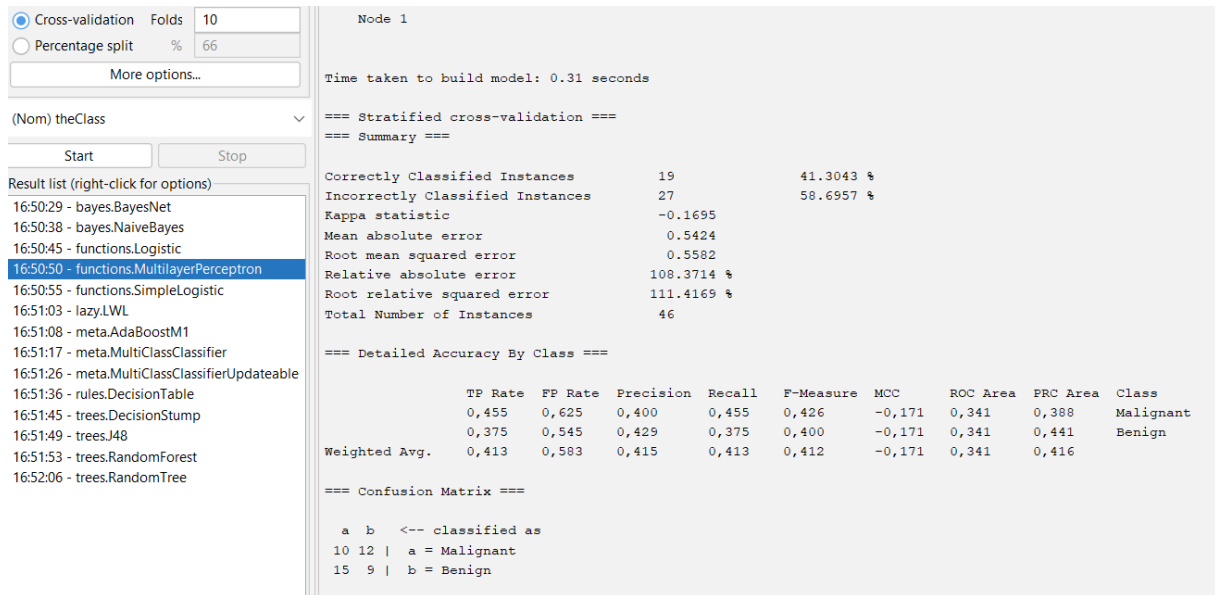


Figure 4.23 : Résultats de classification obtenus par le classifieur Multi Layer Perceptron

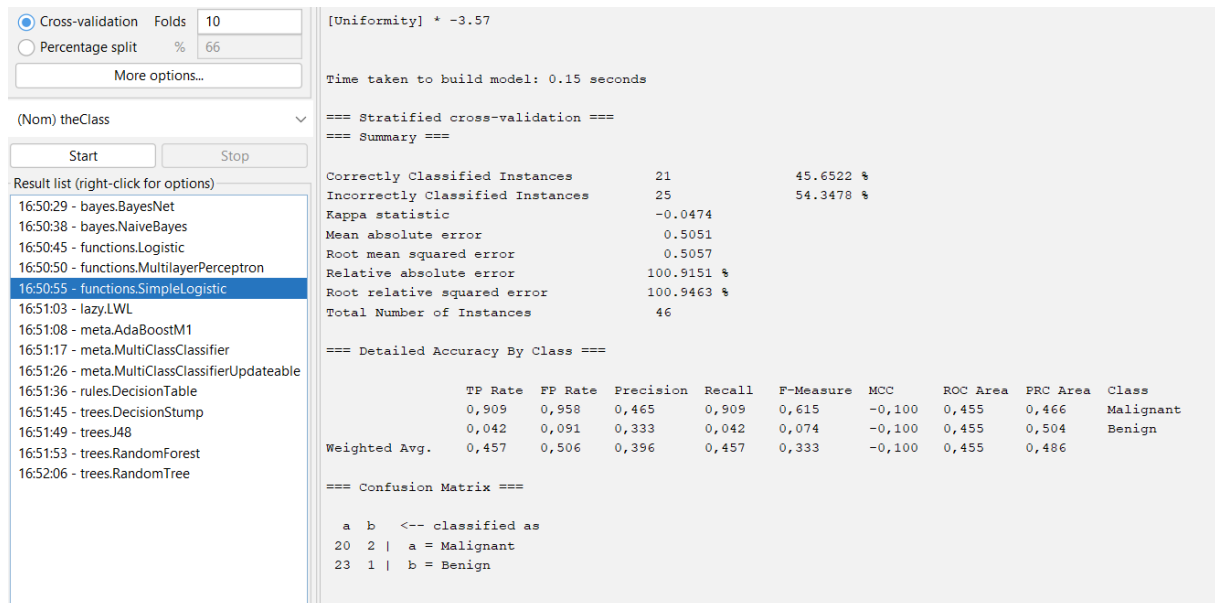


Figure 4.24 : Résultats de classification obtenus par le classifieur Simple Logistic

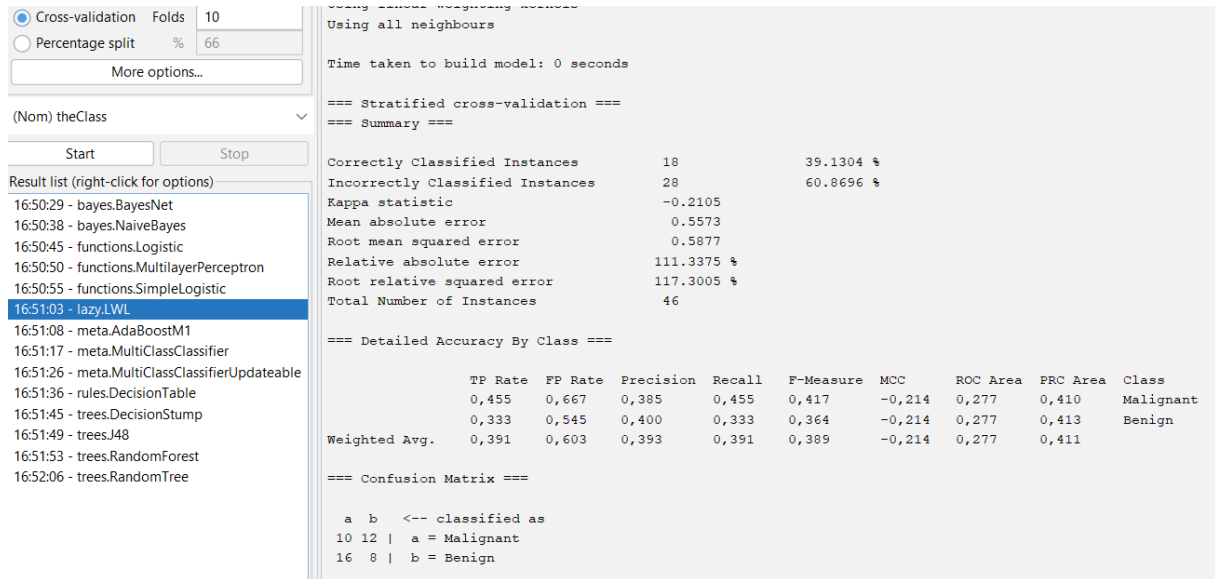


Figure 4.25 : Résultats de classification obtenus par le classifieur LWL

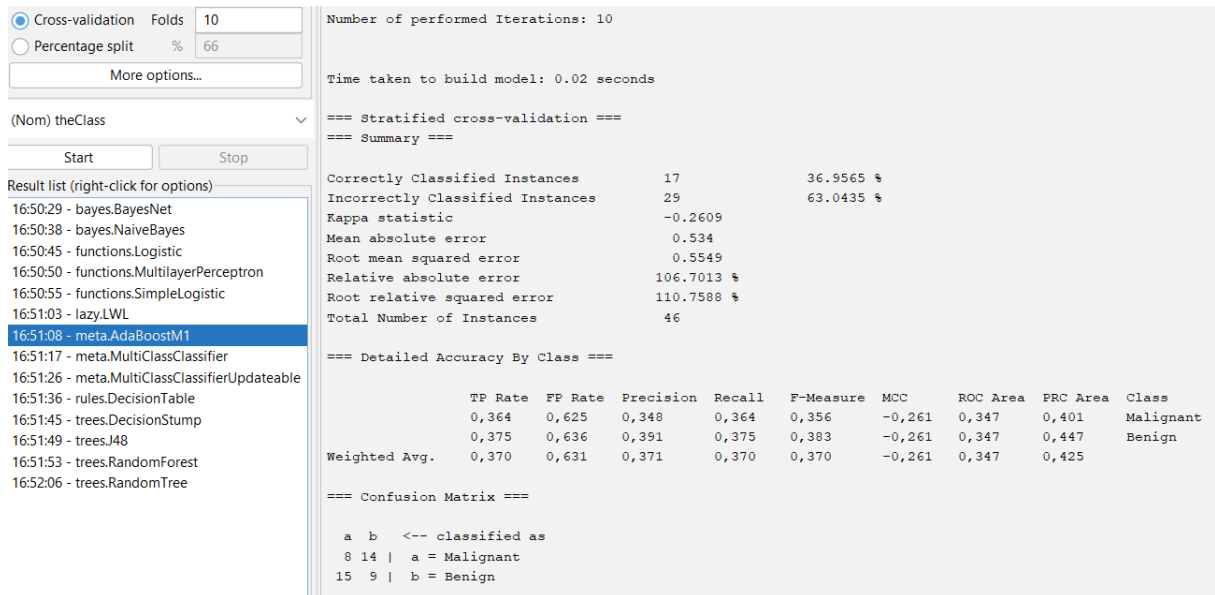


Figure 4.26 : Résultats de classification obtenus par le classifieur Ada Boost M1

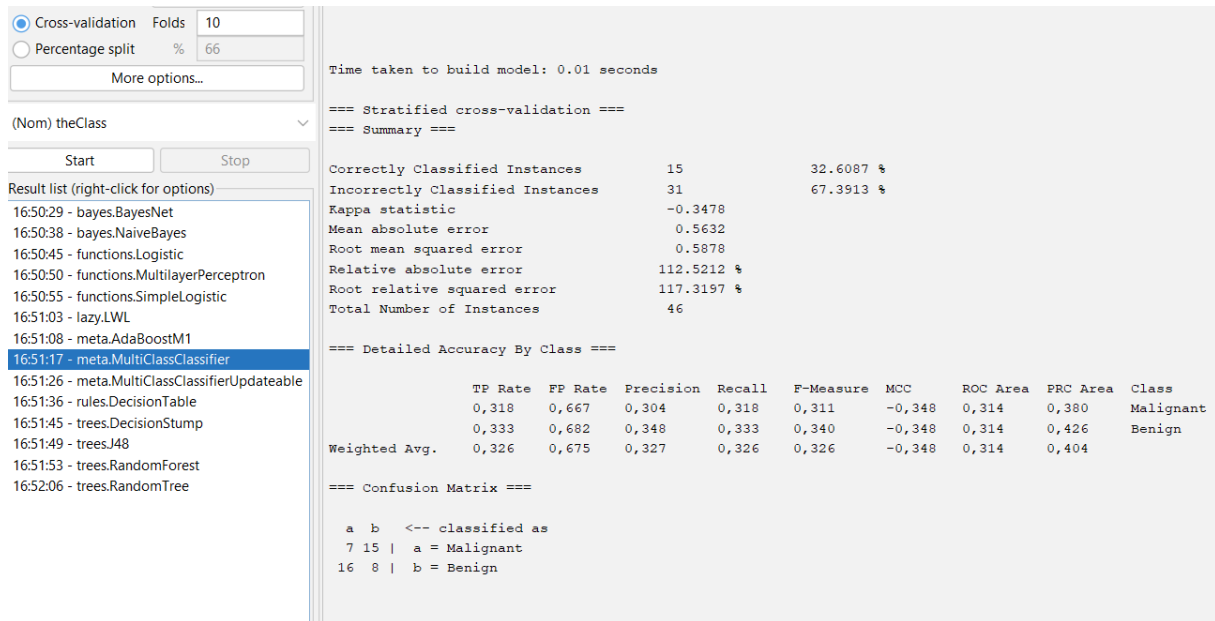


Figure 4.27 : Résultats de classification obtenus par le classifieur Multi Class Classifier

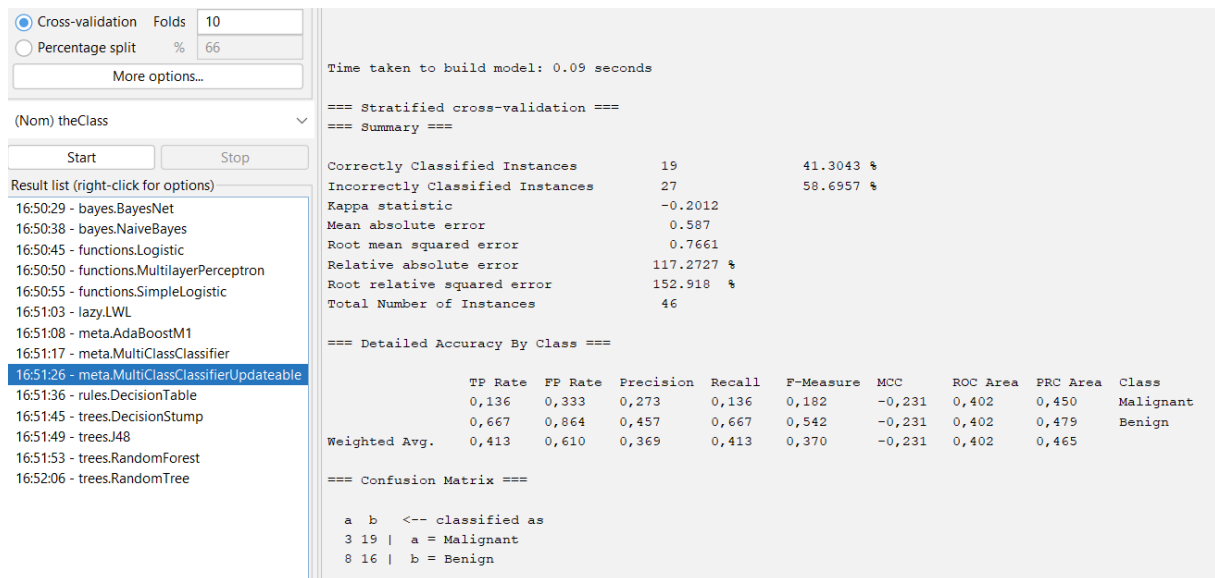


Figure 4.28 : Résultats de classification obtenus par le classifieur Multi Class Classifier Updateable

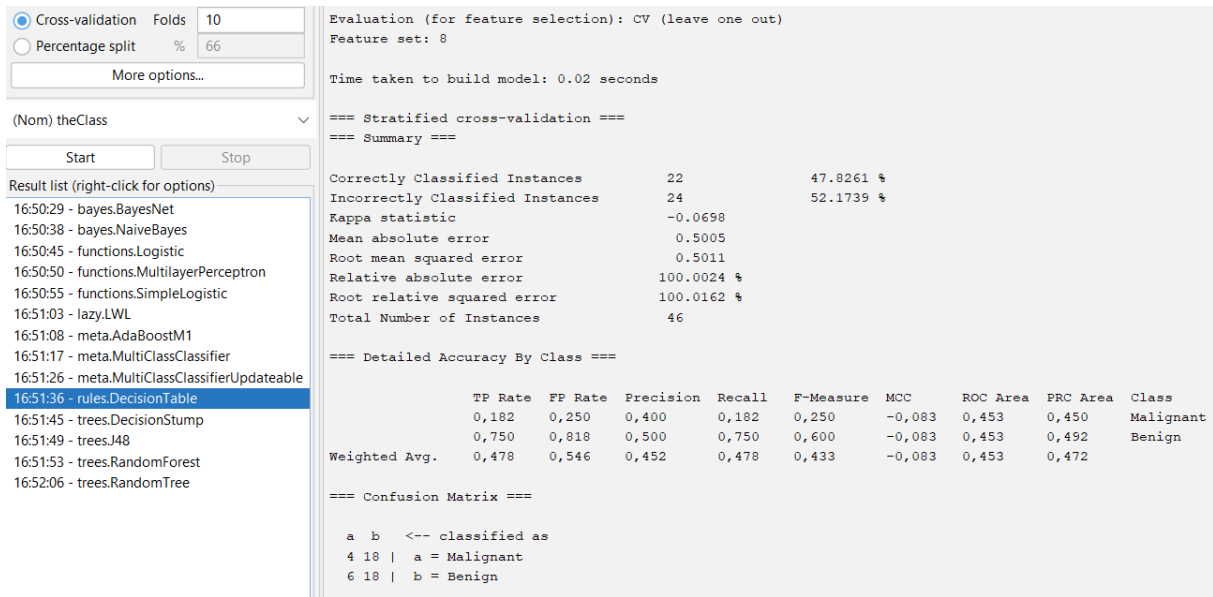


Figure 4.29 : Résultats de classification obtenus par le classifieur Decision Table

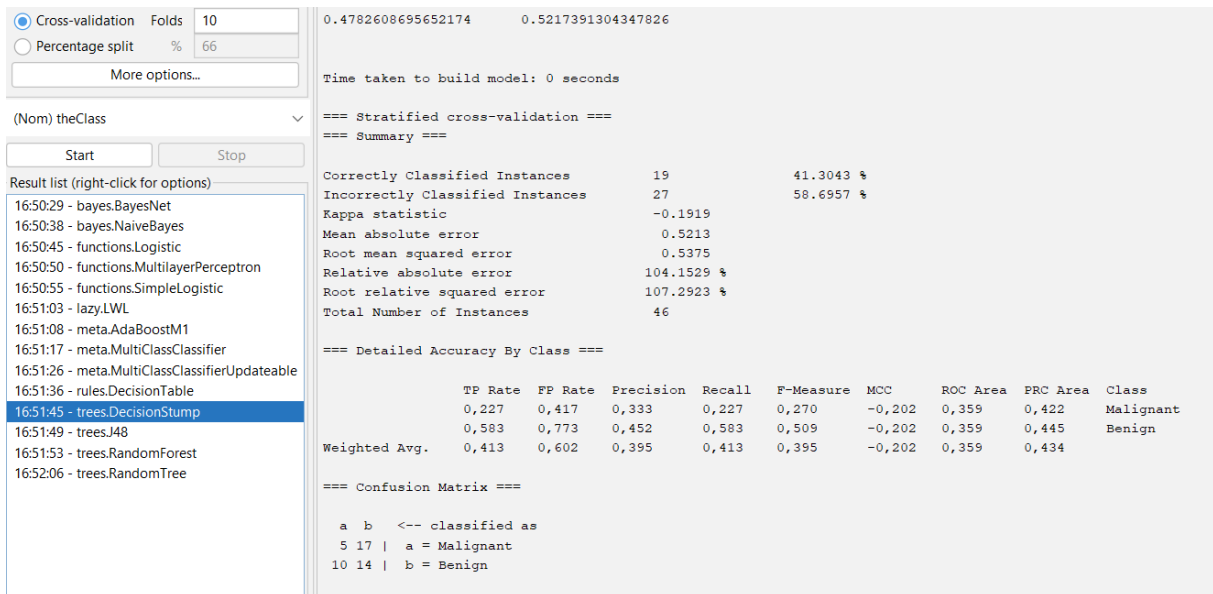


Figure 4.30 : Résultats de classification obtenus par le classifieur Decision Stump

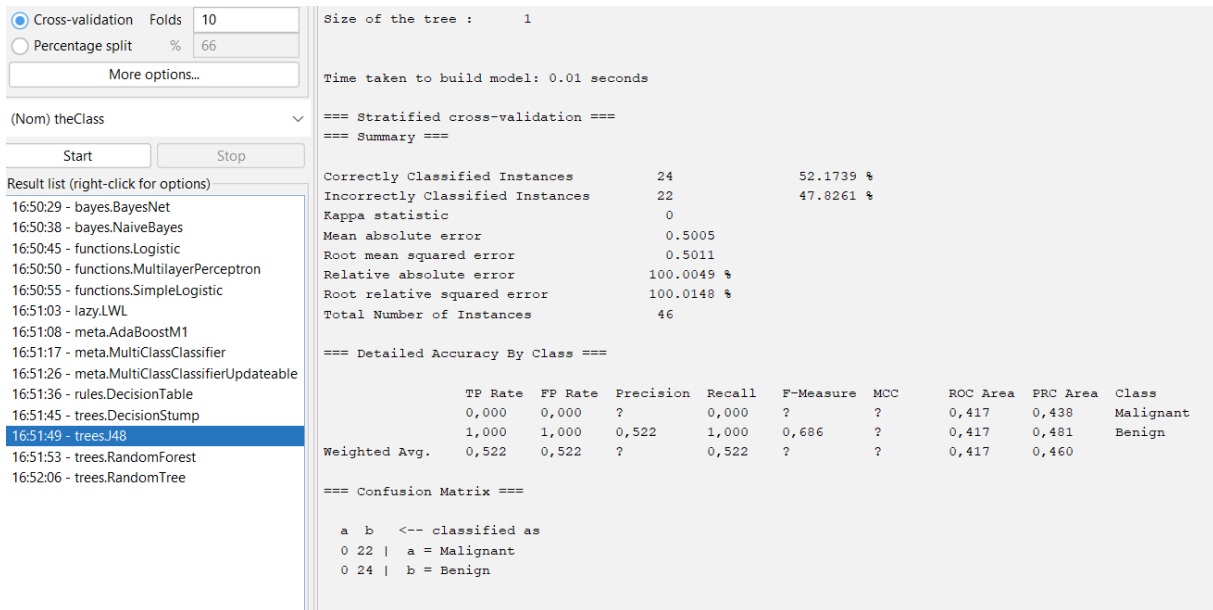


Figure 4.31 : Résultats de classification obtenus par le classifieur J48

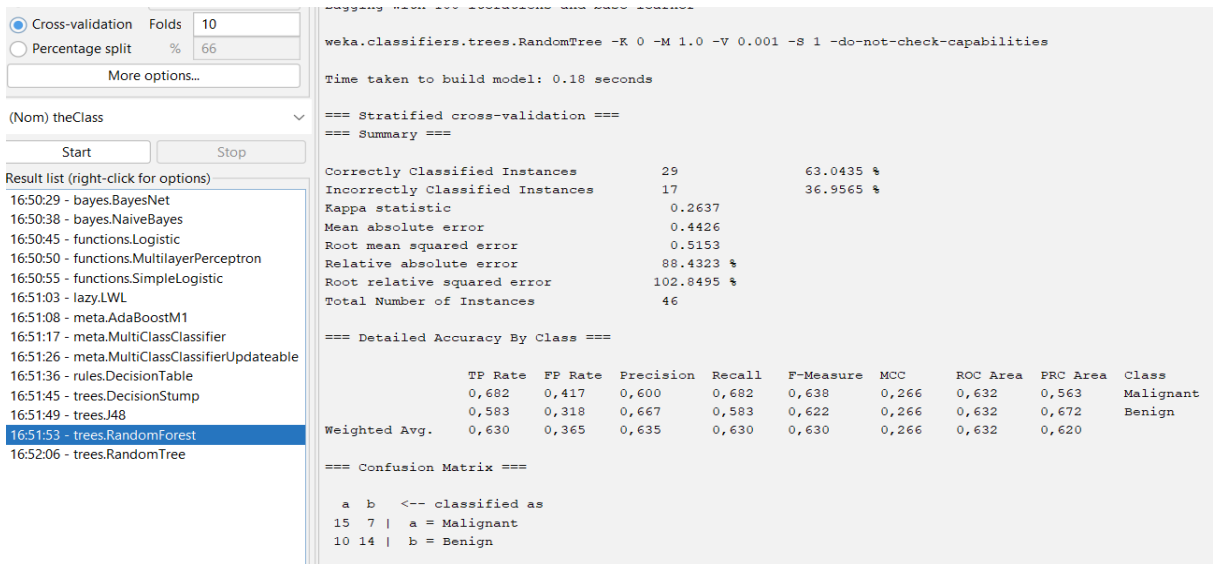


Figure 4.32 : Résultats de classification obtenus par le classifieur Random Forest

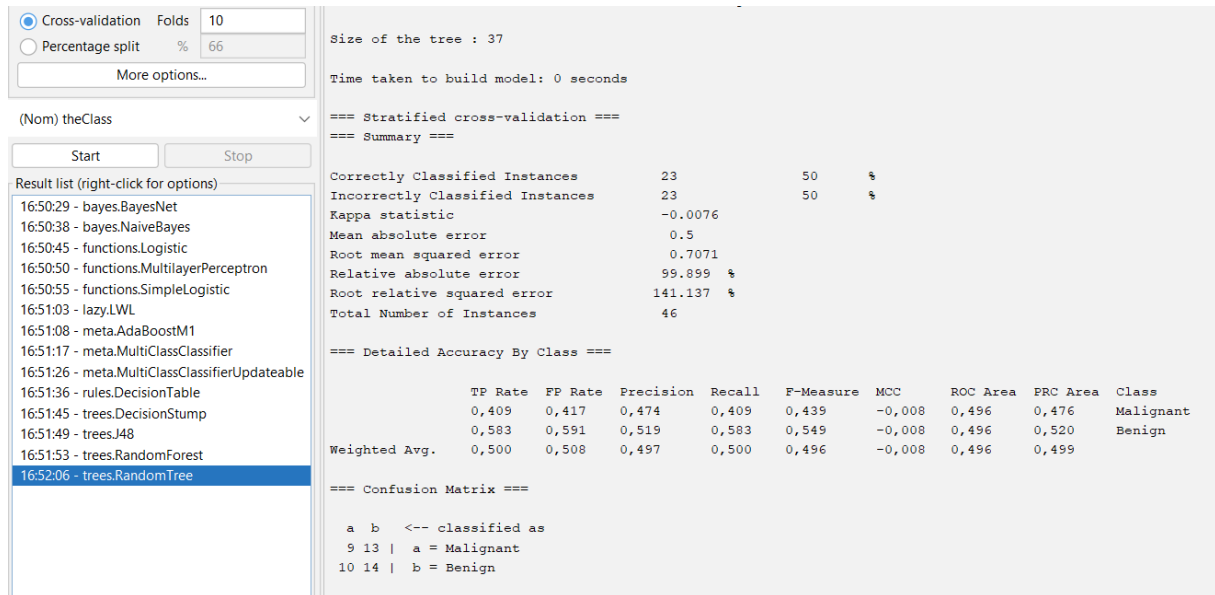


Figure 4.33 : Résultats de classification obtenus par le classifieur Random Tree

Classifiers	Précision	Sensibilité	Spécificité	MC		
				M	B	
				VP	FP	M
				FN	VN	B
BayesNet	52.17%	0%	52.17%	0	22	
				0	24	
NaiveBayes	43.48%	44.12%	41.67%	15	7	
				19	5	
Logistic	32.61%	30.43%	34.78%	7	15	
				16	8	
Multi-Layer-Perception	41.30%	40%	42.86%	10	12	
				15	9	
Simpe Logistic	45.65%	46.51%	33.33%	20	2	
				23	1	
LWL	39.13%	38.46%	40%	10	12	
				16	8	
Ada BoostMI	36.96%	34.78%	39.13%	8	14	
				15	9	
Multi-Class-Classifier	32.61%	30.43%	34.78%	7	15	
				16	8	
Multi-Class-Classifier-UpdateAble	41.30%	27.27%	45.71%	3	19	
				8	16	
Decision Table	47.83%	40%	50%	4	18	
				6	18	
Decision Stump	41.30%	33.33%	45.16%	5	17	
				10	14	
J48	52.17%	0%	52.17%	0	22	
				0	24	
Random Forest	63.04%	60%	66.67%	15	7	
				10	14	
Random Tree	50%	47.37%	51.85%	9	13	
				10	14	

Tableau 4.1 : Récapitulatif des performances des différents classifieurs

Résultats et discussion :

- La segmentation par la méthode k-means est appliquée sur des images mammographiques de la base MIAS. Les résultats obtenus sont satisfaits et clairs.
- D'après les résultats présentés précédemment, nous constatons que :

- le taux de la précision le plus élevé (**63.04%**) est atteint en utilisant le classifieur « **Random Forest** » avec 60% de sensibilité et 66.67% de spécificité.
- Les taux de la précision atteints par les autres classifieurs sont entre 32.61% et 52.17%.
- Le nombre de cas correctement classés comme masse maligne sont élevés en utilisant les classifieurs : NaiveBayes, Simpe Logistic, Random Forest, avec les valeurs de VP = 15 ; 20 et 15 respectivement.
- Les résultats expérimentaux obtenus témoignent aussi que :
 - L'approche proposée est bonne, et elle est capable de distinguer les masses malignes des masses bénignes dans les images mammographiques avec une précision de **63.04%**, car le nombre des images segmentées ne dépassent pas 46.
 - L'utilisation de la méthode LBP et les caractéristiques statistiques de texture influence positivement sur le taux de classification, donc ça nous encourage de rajouter d'autres caractéristiques statistiques dans les prochains travaux, car le bon choix des caractéristiques est l'étape clé qui permet d'obtenir un taux de classification élevé.
 - La plupart des classifieurs donnent des résultats encourageants.

4.4 Conclusion :

Ce système est d'une grande utilité dans le domaine médical surtout pour une telle maladie qui est le cancer du sein. Dans le monde d'aujourd'hui, la robotisation prend de l'ampleur dans tous les domaines donc coté médical, elle permet un gain de temps. Le temps dans la médecine est d'une grande considération l'efficacité et la rapidité tout cela compte pour la survie d'un patient. Donc la conception de tel système est d'une grande utilité pour faciliter le travail du médecin.

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

Nous venons de voir quelques aperçu de notre système, le temps de trouver le tumeur dépend des images car certaines ont le tumeur pas facile à diagnostic tellement minuscule. Donc nous pouvons que le système fonctionne avec un temps d'exécution dépendant du k et du runs et de l'image en question. L'application du LBP donne plus d'avantage à extraire les caractéristiques du premier ordre qui facilite la classification des tumeurs ce qui est la finalité du système.

Conclusions Générales et perspectives :

Le cancer du sein étant la cause du décès de plusieurs femmes sans le diagnostic de la maladie qui se développe au fur des années sans le savoir et même le découvrant tardivement ce serait difficile de sauver la patiente ce qui est d'une grande difficulté. La tumeur doit être diagnostiquée au plus vite afin de pouvoir la retirer avant une certaine évolution ce qui sauverait la vie de milliers de femmes victimes de cette maladie. Alors grâce à ce système le travail du médecin est réduit, il travaille dans l'efficacité et la rapidité afin de diagnostiquer la tumeur au plus vite à ces patients et cela minimiserait le taux de décès provenant de cette maladie.

Nous avons tout d'abord donné un détail général du cancer de sein, en second lieu une description et le fonctionnement des systèmes d'aide assistés par ordinateurs. Ensuite, comment utiliser le k-means dans la segmentation des masses mammaires et enfin la conception de notre système et son implémentation qui nous montrent un peu le déroulement de notre système mise en place

Plusieurs systèmes d'aide au diagnostic assisté par ordinateur ont été conçus jusqu'à nos jours plus précisément pour le cancer du sein. De nos jours ces différents systèmes continuent d'évoluer avec le temps et les expériences de recherche.

Nos perspectives consistent à :

- améliorer notre base d'image, en ajoutant d'autres cas afin de l'enrichir encore plus.
- En outre, utiliser d'autres types de caractéristiques plus pertinentes pour augmenter le taux de classification correcte.
- Faire une comparaison avec d'autres travaux de recherche récents dans le domaine afin de prouver la validité de notre méthode proposée.

Ce système ne serait plus seulement plus que pour le cancer du sein ; on pourrait l'ajouter d'autres fonctionnalités afin qu'il puisse diagnostiquer d'autres cancers ce qui est un avantage dans la médecine. L'avantage est qu'il est adaptable donc facile à modifier.

Références Bibliographiques :

- [1] Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E. R., & Zwiggelaar, R. (2010). "A review of automatic mass detection and segmentation in mammographic images". *Medical Image Analysis*, 14(2), 87–110.
- [2] Sickles, E. A. (1997). "Breast cancer screening outcomes in women ages 40-49 : Clinical experience with service screening using modern mammography". *JNCI Monographs*, 1997(22), 99–104.
- [3] Liu, B., Li, X., Wang, H., Zhao, S., Li, J., Qu, G., & Wang, F. (2020). "Retyping of triple-negative breast cancer based on clustering method". *Expert Systems*, e12583.
- [4] Al-Ayyoub, M., & Al-Zghool, D. (2013). "Determining the type of long bone fractures in x-ray images". *WSEAS Transactions on Information Science and Applications*, 10(8), 261–(1) (1)270.
- [5] Fu, J. C., Lee, S. K., Wong, S. T. C., Yeh, J. Y., Wang, A. H., & Wu, H. K. (2005). "Image segmentation feature selection and pattern classification for mammographic microcalcifications". *Computerized Medical Imaging and Graphics*, 29(6), 419–429.
- [6] Azamjah, N., Soltan-Zadeh, Y., & Zayeri, F. (2019). "Global trend of breast cancer mortality rate : A 25-year study". *Asian Pacific journal of cancer prevention : APJCP*, 20(7), 2015.
- [7] Gaudin, N. (2013). "The International Agency for Research on Cancer". World Health Organisation. ARC Communications, Lyon/Geneva.[8] (Organization, 2020) World Health Organization, 2020
- [9] Roy.A, Singh.B, Banchhor.S, Verma.K, " Segmentation of malignant tumours in mammogram images : A hybrid approach using convolutional neural networks and connected component analysis ", pp 1-21, DOI : 10.1111/exsy.12826
- [10] World Health Organization. Accessed : Nov. 19, 2020. [Online]. Available :<https://www.who.int/news-room/fact-sheets/detail/cancer>
- [11] TOZ.G, (Member, IEEE), AND PAKIZE ERDO.MU, " A Novel Hybrid Image Segmentation Method for Detection of Suspicious Regions in Mammograms Based on Adaptive Multi-Thresholding (HCOW) ", IEEE ACCESS, pp 1-16, Digital Object Identifier 10.1109/ACCESS.2021.3089077
- [12] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha, "Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform," *Measurement*, vol. 146, pp. 800–805, 2019.
- [13] D. Muduli, R. Dash, and B. Majhi, "Automated breast cancer detection in digital mammograms : a moth flame optimization based ELM approach," *Biomedical Signal Processing and Control*, vol. 59, article 101912, 2020.

- [14] S. Sapate, S. Talbar, A. Mahajan, N. Sable, S. Desai, and M. Thakur, “Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 290–305, 2020
- [15] Edité par Élisabeth.L, Leichtnam-Dugarin.L, Emmanuel.A, Jean-Marc.C, Nadine.D, René.G, Pierre.K, Louis.M, Anne.T, Philippe.V, Sylvie.B, Julien.C, Marie.D, Valérie.D, Béatrice. F, Thierry.P, T. Bachelot, A. Brémond, E. Delay, F. Demma, J.-M. Dilhuydy, M.-H. Dilhuydy, B. Fontanière, H. Hoarau, N. Moumjid-Ferdjaoui, M. Porché, P. Saltel, P. Thiesse, M. Véron, “ Comprendre le cancer du sein”, en collaboration avec L’INSTITUT NATIONAL DU CANCER, mis à jour 2022, pp 1-116
- [16] Elżbieta.K; Fatima.C; Jean-Yves.D; Claire.B; Francesca.L; and Svetlana.J, Tanja.S, Deborah.F and Anita.M, “ Breast Cancer ‘’, written by Kstorfin Medical Communications Ltd on behalf of ESMO, pp 1-70
- [17] (Web 01) Créé par Docteur Caroline, ”Anatomie du sein”, Dernière mise en jour 07/01/2019, consulté le 07/05/2022, Disponible sur : <https://www.docteurclic.com/maladie/mastite.aspx>
- [18] (Web 02) Consulté le 06/05/2022, Disponible sur : <https://acteurdemasante.lu/fr/cancer-du-sein/le-sein-a-la-decouverte-de-son-anatomie/>
- [19] (web 03) Consulté le 07/05/2022, “Cellules cancéreuses”, Disponible sur : <http://nuxeo.edel.univ-poitiers.fr/nuxeo/site/esupversions/5de6eb92-1d05-414d-9553-159f5ee6dfd2>
- [20] (Web 04) Consulté le 07/05/2022, “Cancer canalaire infiltrant et cancer canalaire in situ”, Disponible sur : <http://kelly-en-2010.over-blog.com/article-le-22-fevrier-47673733.html>
- [21] (Web 05) Consulté le 07/05/2022, “ Les quatre secteurs du sein ‘’, Disponible sur : <https://www.futura-sciences.com/sante/dossiers/medecine-cancer-sein-1154/page/3/>
- [22] (Web 06) Consulté le 07/05/2022, “ Chirurgie conservatrice ‘’, disponible sur : <https://www.futura-sciences.com/sante/dossiers/medecine-cancer-sein-1154/page/4/>
- [23] (Web 07) Consulté le 07/05/2022, “ Récepteurs hormonaux d’un cancer hormonosensible ‘’, Disponible sur : <https://www.revmed.ch/revue-medicale-suisse/2013/revue-medicale-suisse-387/hormonotherapie-dans-le-cancer-du-sein-efficacite-et-effets-adverses>
- [24] Das S., Abraham A., Konar A. “Automatic clustering using an improved differential evolution algorithm”. *IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans*, 2008 ; 38 : 218-237.
- [25] Shokrgozar N., Sobhani FM. “Customer segmentation of bank based on iscovering of their transactional relation by using data mining algorithms”. *Modern Applied Science* 2016 ; 10(10) : 283-286.

- [26] Khan SS., Ahmad A. Cluster centre initialization algorithm for k-means cluster. In Pattern Recognition Letters 2004 ; 1293–1302.
- [27] Kaur A. ‘‘Comparative analysis of segmentation algorithms for brain tumor detection in MR images’’. Medicine 2017.
- [28] Yang MS., Sinaga KP. ‘‘A feature-reduction multi-view K-means clustering algorithm’’. IEEE 2019 ; 7 : 114472-114486.
- [29] Lin H., Ji Z. ‘‘Breast cancer prediction based on K-Means and SOM Hybrid Algorithm’’. In Journal of Physics : Conference Series. IOP Publishing 2020 ;1624(4) : 1-7.
- [30] Aswathy MA., Jagannath M. ‘‘Performance analysis of segmentation algorithms for the detection of breast cancer’’. Procedia Computer Science 2020 ; 167 : 666-676.
- [31] Çiklaçandır FGY., Ertaylan A., Bınzat U., Kut A. ‘‘Lesion detection from the ultrasound images using k-means algorithm’’. Medical Technologies Congress (TIPTEKNO) 2019 ; 1-4.
- [32] Bottou L., Lin CJ.’’ Support vector machine solvers’’. Large Scale Kernel 2007 ; 3(1) : 301–320.
- [33] Kuşcu AC., Erol H. ‘‘Diagnosis of Breast Cancer by K-Mean Clustering and Otsu Thresholding Segmentation Methods’’. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2022 ; 5(1) :258-281.
- [34] Mahmood, T., Li, J., Pei, Y., Akhtar, F., Rehman, M.U. and Wasti, S.H. (2022) ‘Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach’. PLoS ONE, vol. 17, <https://doi.org/10.1371/journal.pone.0263126>.
- [35] Astley, S. M. and Gilbert, F. J. (2004). ‘‘Computer-aided detection in mammography’’. Clinical Radiology, 59(5) :390–399.
- [36] Hadjiiski, L., Chan, H. P., Sahiner, B., Helvie, M. A., Roubidoux, M. A., Blane, C., Paramagul, C., Petrick, M. N., Bailey, J., Klein, K., Foster, M., Patterson, S., Adler, A., Nees, A., and Shen, J. (2004).’’ Improvement in radiologists characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis : An roc study’’. Radiology, pages 255–265.
- [37] (Web 08) Consulté le 13/05/2022, ‘‘ Computer-aided diagnosis ‘’, Disponible sur : https://en.wikipedia.org/wiki/Computer-aided_diagnosis
- [38] Bird, R. E., Wallace, T. W., and Yankaskas, B. C. (1992). ‘‘ Analysis of cancers missed at screening mammography. Radiology ‘’, 184(3) :613–617.
- [39] IMENE CHEIKHROUHOU Epse KACHOURI, ‘‘ Description et classification des masses mammaires pour le diagnostic du cancer du sein ‘’, Thèse Doctorat, UNIVERSITE D’EVRY VAL D’ESSONNE, pp 1-160, 2012.
- [40] Ciatto, S., Turco, M. R. D., Risso, G., Catarzi, S., Bonaldi, R., Viterbo, V., Gnutti, P., Guglielmoni, B., Pinelli, L., Pandiscia, A., Navarra, F., Lauria, A., Palmiero, R., and Indovina,

- P. L. (2003). "Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography". *European journal of radiology*, 45(2) :135–138.
- [41] Krupinski, E. A. and Nishikawa, R.M. (1997). "Comparison of eye position versus computer identified microcalcification clusters on mammograms". *Medical Physics*, 24(1) :17–23.
- [42] Dengler, J., Behrens, S., and Desaga, J. F. (1993). "Segmentation of microcalcifications in mammograms". *IEEE Transactions on Medical Imaging*, 12(4) :634–642.
- [43] Chan, H. P., Vyborny, C. J., MacMahon, H., Metz, C. E., Doi, K., and Sickles, E. A. (1987). "Digital mammography : Roc studies of the effects of pixel size and unsharpmask filtering on the detection of subtle microcalcifications". *Investigative Radiology*, 22(7) :581–589.
- [44] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Geer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. (1987). "Adaptive histogram equalization and its variations". *Computer Vision, Graphics, and Image Processing*, 39(3) :355–368.
- [45] Ji, T. L., Sundareshan, M. K., and Roehrig, H. (1994). "Adaptive image contrast enhancement based on human visual prospérités". *IEEE Transaction on Medical Imaging*, 13(4) :573–586.
- [46] Kim, J. K., Park, J. M., Song, K. S., and Park, H. W. (1997). "Adaptive mammographic image enhancement using first derivative and local statistics". *IEEE Transactions on Medical Imaging*, 16(5) :495–502.
- [47] Morrow, W. M., Paranjape, R. B., Rangayyan, R. M., and Desautels, J. E. L. (1992)." Region-based contrast enhancement of mammograms". *IEEE Transactions on Medical Imaging*, 11(3) :392–406.
- [48] Gordon, R. and Rangayyan, R. M. (1984). "Feature enhancement of film mammograms using fixed and adaptive neighborhoods". *Applied Optics*, 23(4) :560–564.
- [49] Rangayyan, R. M. and Nguyen, H. N. (1987)." Pixel-independent image processing techniques for noise removal and feature enhancement". In *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*.
- [50] Laine, A., Fan, J., and Yan, W. H. (1995)." Wavelets for contrast enhancement of digital mammography. *IEEE Engineering in Medicine and Biology Magazine*", 14(5):536–550.
- [51] Sivaramakrishna, R., Obuchowski, N. A., Chilcote, W. A., Cardenosa, G., and Powell, K. A. (2000). "Comparing the performance of mammographic enhancement algorithms: a preference study". *American Journal of Roentgenology*, 175:45–51.
- [52] Matsubara, T., Fujita, H., Hara, T., Kasai, S., Otsuka, O., Hatanaka, Y., and Endo, T. (1998)." Development of a new algorithm for detection of mammographic masses". In *International Workshop on Digital Mammography*, pages 139–142.

- [53] Matsubara, T., Fujita, H., Kasai, S., Goto, M., Tani, Y., Hara, T., and Endo, T. (1997). ‘‘Development of new schemes for detection and analysis of mammographic masses’’. In International Conference on Intelligent Information Systems, pages 63–66.
- [54] Mudigonda, N. R., Rangayyan, R. M., and Desautels, J. E. L. (2001). ‘‘Detection of breast masses in mammograms by density slicing and texture flow-field analysis’’. IEEE Transactions on Medical Imaging, 20(12):1215–1227.
- [55] Roberts, L. G. (1963). ‘‘Machine perception of three-dimensional solids. Optical and Electrooptical Information processing’’, pages 159–197.
- [56] Sobel, I. E. (1970). ‘‘Camera models and machine perception’’. Thèse de Doctorat, Stanford University.
- [57] Canny, J. (1986). ‘‘A computational approach to edge detection’’. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679–698.
- [58] Prewitt, J. M. S. (1970). Object enhancement and extraction. Picture Processing and Psychopictorics. Academic Press, New York.
- [59] Web 09, consulté le 04/10/2022, https://media.springernature.com/lw685/springer-static/image/chp%3A10.1007%2F978-3-540-78450-0_6/MediaObjects/978-3-540-78450-0_6_Fig1_HTML.jpg
- [60] Chokri FERKOUS, ‘‘Extraction des connaissances pour la segmentation d’images mammographiques’’, Thèse Doctorat, 2017, pp :1-171, <https://www.researchgate.net/publication/338104120>
- [61] Haralick, R. M., & Shanmugam, K. (1973). ‘‘Textural features for image Classification’’. IEEE Transactions on systems, man, and cybernetics, 3 (6), 610-621.
- [62] Ojala T, Pietikäinen M & Mäenpää T (2002b) ‘‘Multiresolution gray-scale and rotation invariant texture classification with local binary patterns’’. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7) : 971–987.
- [63] Pietikä Pietikainen, M., Hadid, A., Zhao, G., Ahonen, T. : ‘‘Computer Vision Using Local Binary Patterns’’. Springer, London (2011)
- [64] Ojala, T., Pietikäinen, M., Harwood, D. : ‘‘A comparative study of texture measures with classification based on feature distributions’’. Pattern Recognit. **29**(1), 51–59 (1996)
- [65] Ojala, T., Pietikainen, M., Harwood, D. : ‘‘Performance evaluation of texture measures with classification based on kullback discrimination of distributions’’. In : 12th IAPR International Conference (1994).
- [66] Ojala, T., Pietikäinen, M., Mäenpää, T. : Gray scale and rotation invariant texture classification with local binary patterns. In : European Conference on Computer Vision. Lecture Notes in Computer Science, vol. 1842, pp. 404–420. Springer, Berlin (2000)

- [67] Huo, Z., Giger, M. L., Vyborny, C. J., Bick, U., Lu, P., Wolverton, D. E., and Shmidt, R. A. (1995). ‘‘Analysis of spiculation in the computerized classification of mammographic masses’’. *Medical Physics*, 22(10):1569–1579.
- [68] Kupinski, M. A. and Giger, M. L. (1998). ‘‘Automated seeded lesion segmentation on digital mammograms’’. *IEEE Transactions on Medical Imaging*, 17(4).
- [69] Cao, Y., Hao, X., and Xia, S. (2009). ‘‘An improved region-growing algorithm for mammographic mass segmentation’’. In *Medical imaging, parallel processing of images, and optimization techniques.*, volume 7497.
- [70] Horowitz, S. L. and Pavlidis, S. (1974). ‘‘Picture segmentation by a directed split and merge procedure’’. In *Second International Joint Conference on Pattern Recognition.*, pages 424–433.
- [71] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E. L., and Alim, O. A. (1997). ‘‘Measures of acutance and shape for classification of breast tumors’’. *IEEE Transactions on Medical Imaging*, 16(6):799–810.
- [72] Kachouri, R., Djemal, K., and Maaref, H. (2010). ‘‘Multi-model classification method in heterogeneous image databases’’. *Pattern Recognition*, 43(12):4077–4088.
- [73] Sajida Parveen, Nadeem Naeem and Jherna Devi, ‘‘REVIEW ON LOCAL BINARY PATTERN (LBP) TEXTURE DESCRIPTOR AND ITS VARIANTS.’’, *International Journal Of Advanced Research (IJAR)*, Journal Homepage : - www.journalijar.com, Article DOI : 10.21474/IJAR01/4169, DOI URL : <http://dx.doi.org/10.21474/IJAR01/4169>, pp 1-10
- [74] Choquet, G. (1978). *Cours de topologie.*, volume 54. Dunod, 2^e édition.
- [75] McQueen, J. (1967). ‘‘Some methods for classification and analysis of multivariate observations’’. In Cam, L. and Neyman, J., editors, ‘‘Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability’’.
- [76] Kohonen, T. (1984). ‘‘Self-Organization and Associative Memory’’. Berlin : Springer-Verlag.
- [77] Vapnik, N. (1999). ‘‘An overview of statistical learning theory’’. *IEEE Trans Neural Networks*, 10 :988–999.
- [78] (Web 09) Consulté le 15/05/2022, ‘‘Steps for performing clustering analysis’’, Disponible sur :<https://www.semanticscholar.org/paper/HEALTHCARE-DATASET-USING-SILHOUETTE-SCORE-VALUE-Ogbuabor-Ugwoke/199d5cb0e1f29360b396628ddc1c63c7f6256765>
- [79] Andika Elok Amalia, Gregorius Airlangga, Affandi Nur Aziz Thohari, ‘‘Breast Cancer Image Segmentation Using K-Means Clustering Based on GPU Cuda Parallel Computing’’, *JURNAL INFOTEL*, Website Jurnal : <http://ejournal.st3telkom.ac.id/index.php/infotel> ISSN : 2085-3688, 2018, pp 1-6; e-ISSN : 2460-0997

- [80] H.S. RAHLI and N. BENAMRANE, “ Interprétation des images mammographiques par la méthode K-means et Search. Harmony ”, Conference Paper pp 1-11 , May 2012, <https://www.researchgate.net/publication/257308096>
- [81] Na, S., Xumin, L., & Yong, G. (2010, April). “Research on k-means clustering algorithm: An improved k-means clustering algorithm”. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on* (pp. 63-67). IEEE.
- [82] (Web 10) Consulté le 15/05/2022, “ Flowchart K-Means Clustering ”, Disponible sur :<http://ba-finance-2013.blogspot.com/2012/09/k-means-clustering.html>
- [83] Younes Benzaki, “ Tout ce que vous voulez savoir sur l’algorithme K-Means’ ”, Mr.Mint Machine Learning, 10 avril 2018, <https://mrmint.fr>
- [84] Fauser, S.; Schwenker, F. “Clustering large datasets with kernel methods. In: *Proceedings of the 21st International Conference on Pattern Recognition*”, 501– 504, 2012.
- [85] Razavi Zadegan, S. M.; Mirzaie, M.; Sadoughi, F. Ranked” *k*-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets”. *Knowledge-Based Systems* Vol. 39, 133–143, 2013.
- [86] Taye Girma Debelee, Friedhelm Schwenker, Samuel Rahimeto, and Dereje Yohannes, “ Evaluation of modified adaptive k-means segmentation algorithm”, *Computational Visual Media*, Vol. 5, No. 4, December 2019, 347–361, <https://doi.org/10.1007/s41095-019-0151-2>
- [87] K.Sushmithawathi and Prof.P.Indra, “ AN AUTOMATED COMPUTER AIDED SEGMENTATION OF SKIN CANCER ”, *International Research Journal of Modernization in Engineering Technology and Science*, Volume:03/Issue:09/September-2021
- [88] J. C. Bezdek, "Objective Function Clustering", *Pattern Recognition with Fuzzy Objective Function Algorithms*, ed : Springer, pp. 43-93, 1981.
- [99] Z. Wang, Q. Song, Y. C. Soh, and K. Sim, "An adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation", *Computer Vision and Image Understanding*, Vol. 117, pp. 1412-1420, 2013.
- [90] S. Kannan, R. Devi, S. Ramathilagam, and K. Takezawa, "Effective FCM noise clustering algorithms in medical images", *Computers in Biology and Medicine*, Vol. 43, pp. 73-83, 2013.
- [91] Jack Sparrow, "Différents types d’algorithmes de clustering", Acervo Lima, 5 juillet 2022, consulté le 04/10/2022, <https://fr.acervolima.com/differents-types-d-algorithmes-de-clustering/>
- [92] Z.-Z. Han, H.-J. Chen, J.-P. Li, and C. Yao, "A novel method of mass segmentation in mammogram", In: *Proc. of the 2012 International Conference on Systems and Informatics*, pp. 1412-1416, 2012.

- [93] L. Vivona, D. Cascio, F. Fauci, and G. Raso, "Fuzzy technique for microcalcifications clustering in digital mammograms", *BMC Medical Imaging*, Vol. 14, p. 23, 2014.
- [94] A. Touil and K. Kalti, "Iterative fuzzy segmentation for an accurate delimitation of the breast region", *Computer Methods and Programs in Biomedicine*, Vol. 132, pp. 137-147, 2016.
- [95] Andika Elok Amalia, Gregorius Airlangga, Affandi Nur Aziz Thohari, ‘‘ Breast Cancer Image Segmentation Using K-Means Clustering Based on GPU Cuda Parallel Computing ‘’, *JURNAL INFOTEL*, Website Jurnal : <http://ejournal.st3telkom.ac.id/index.php/infotel> ISSN : 2085-3688, 2018, pp 1-6; e-ISSN : 2460-0997
- [96] Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D., Piñeros, M., Znaor, A. and Bray, F. (2019) „Estimating the global cancer incidence and mortality in 2018 : GLOBOCAN sources and methods“, *International Journal of Cancer*, vol. 144, pp. 1941-1953, doi : 10.1002/ijc.31937.
- [97] Mahmood, T., Li, J., Pei, Y., Akhtar, F., Rehman, M.U. and Wasti, S.H. (2022) ‘Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach’. *PLoS ONE*, vol. 17, <https://doi.org/10.1371/journal.pone.0263126>.
- [98] Oliveira, F.S.S., Filho, A.O.C., Silva, A.C., Paiva, A.C., Gattass, M. (2014) „Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM“, *Computers in Biology and Medicine*, vol. 57, pp. 42-53.
- [99] Lévy, D. and Jain, A. (2016) ‘Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks’, *arXiv Preprint*. arXiv : 1612.00542.
- [100] Ahmed, L, Iqbal, MM, Aldabbas, H, Khalid, S, Saleem, Y, Saeed, S. (2020) ‘Images data practices for Semantic Segmentation of Breast Cancer using Deep Neural Network’, *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, <https://doi.org/10.1007/s12652-020-01680-1>.
- [101] Jiao, Z., Gao, X., Wang, Y. and Li, J. (2016) ‘A deep feature based framework for breast masses classification’. *Neurocomputing*, vol. 197, pp. 221–231. <https://doi.org/10.1016/j.neucom.2016.02.060>
- [102] V. S. Gnanasekaran, S. Joypaul, P. Meenakshi Sundaram, and D. D. Chairman, “Deep learning algorithm for breast masses classification in mammograms,” *IET Image Processing*, vol. 14, no. 12, pp. 2860–2868, 2020.
- [103] B. Swiderski, P. O. Gielata, P. Olszewski, S. Osowski, and M. Kołodziej, “Deep neural system for supporting tumor recognition of mammograms using modified GAN,” *Expert Systems with Applications*, vol. 164, article 113968, 2021.
- [104] R. Agarwal, O. Díaz, M. H. Yap, X. Llado, and R. Martí, “Deep learning for mass detection in full field digital mammograms,” *Computers in biology and medicine*, vol. 121, article 103774, 2020.

- [105] L. G. Falconi, M. Perez, W. G. Aguila, and A. Conci, “Transfer learning and fine tuning in breast mammogram abnormalities classification on cbis-ddsm database,” *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 2, pp. 154–165, 2020.
- [106] M. Amalmary and A. Prakash, Elephant herding optimization with bi-directional long shortterm memory network (ehobilstm) for breast cancer detection in digital mammograms.
- [107] A. M. Alqudah, H. M. S. Algharib, A. M. S. Algharib, and H. M. S. Algharib, “Computer aided diagnosis system for automatic two stages classification of breast mass in digital mammogram images,” *Biomedical Engineering: Applications, Basis and Communications*, vol. 31, no. 1, article 1950007, 2019.
- [108] N. S. Ismail and C. Sovuthy, “Breast cancer detection based on deep learning technique,” in *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)*, pp. 89–92, Kuching, Malaysia, 2019.
- [109] R. Touahri, N. AzizI, N. E. Hammami, M. Aldwairi, and F. Benaida, “Automated breast tumor diagnosis using local binary patterns (lbp) based on deep learning classification,” in *2019 International Conference on Computer and Information Sciences (ICCIS)*, pp. 1–5, Sakaka, Saudi Arabia, 2019.
- [110] Aqsa Mohiyuddin, Asma Basharat, Usman Ghani, Vesely Peter , Sidra Abbas , Osama Bin Naeem,5 and Muhammad Rizwan, “ Breast Tumor Detection and Classification in Mammogram Images Using Modified YOLOv5 Network “, *Hindawi Computational and Mathematical Methods in Medicine Volume 2022*, Article ID 1359019, 16 pages <https://doi.org/10.1155/2022/1359019>
- [111] Web 11, consulté le 22/09/2022 ; disponible : <https://deptinfo.labri.fr/ENSEIGNEMENT/programmation2/intro-eclipse/>
- [112] Web 12 ; consulté le 22/09/2022, <https://www.mammoimage.org/databases/>
- [113] Web 13 ; consulté le 22/09/2022 , <https://datascientest.com/matrice-de-confusion>