

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



UNIVERSITE ECHAHID HAMMA LAKHDAR - EL OUED



FACULTÉ DES SCIENCES EXACTES

Département D'Informatique

Mémoire de Fin D'étude

Présenté pour l'obtention du Diplôme de

MASTER ACADEMIQUE

Domaine : Mathématique et Informatique

Filière : Informatique

Spécialité : Systèmes Distribués et Intelligence Artificielle (SDIA)

Présenté par :

Mehellou Souhaila

Kertiou Fatma Zohra

Thème

Analyse de la texture des images mammaires par une fusion des lois de Zipf et des LBP pour la classification des tumeurs mammaires via les algorithmes génétiques

Soutenue le 12-09- 2019 Devant le jury:

M. Yakoub Mohammed Amine	MCA	Président
M. Guaia sana	MAA	Rapporteur
Dr. Hamoud Meriem	MAA	Encadreur

Année Universitaire: 2019/2020

Remerciements

*Je remercie tout d'abord notre Dieu qui nous a donné la force et la
Volonté pour élaborer ce travail.*

*J'adresse mes vifs remerciements à mon encadreur
Dr. Meriem HAMOUD*

*Qui m'a aidés durant mon travail et par sa patience et ses
Précieux conseils dont Il nous a entourés.
Sans son aide, mon travail n'aurait pas vu la lumière.*

*Je remercie vivement les membres du jury qui m'ont fait l'honneur
D'accepter de juger notre travail.*

*Notre reconnaissance va aussi à tous ceux qui ont collaboré à notre
Formation en particulier les enseignants du département
D'Informatique, de l'université
Echahid hamma lakhdar_El oued*

*Je remercie également tous ceux qui ont participé de près
Ou de loin à élaborer ce travail*

.

Table des matières

Sommaire

Table des matières	II
Liste des figures	IV
Liste des tableaux	VI
Liste des acronymes	VII
Introduction générale	1
I.1. Introduction	2
I.2. Définition de l'analyse d'image	2
I.2.1. Catégories de l'analyse d'images	3
L'analyse d'images regroupe plusieurs catégories [4] [6] citons:.....	3
I.3. Définition de la vision par ordinateur	4
I.4. Concept de l'analyse de la texture d'une image	4
I.4.1. Désignation du terme « Texture »	5
I.4.2. Les méthodes d'analyse de la texture	5
I.4.2.1. Méthodes structurelles	5
I.4.2.2. Méthodes statistiques	7
I.4.2.3. Méthodes spectrales	7
a) Filtres de domaine spatial.....	7
b) Filtres de domaine de fréquence.....	8
I.4.2.4. Méthodes basées sur un modèle	8
I.5. Conclusion	9
II.1. Introduction	11
II.2. Lois de puissance	11
II.3. Les principales de lois puissance	12
II.3.1. Loi de Pareto	12
II.3.2. La loi de Zipf	12
II.3.3. La loi de Zipf inverse	12
II.4. Analyse d'images par les lois de Zipf et de Zipf inverse	13
II.4.1. Analyse d'image par la loi de Zipf	13
II.4.2. Analyse d'image par la loi de Zipf inverse	13
II.4.3. Codages nécessaires à l'analyse d'image par les lois de Zipf et de Zipf inverse	14
II.4.4. Construction des courbes de Zipf et de Zipf inverse	15

a) Construction de la courbe de Zipf.....	15
b) Construction de la courbe de Zipf inverse.....	17
II.5. Conclusion.....	18
III.1. Introduction.....	20
III.2. Le cancer du sein.....	20
III.3. Les types du cancer du sein.....	22
III.3.1. Cancer non invasif.....	22
III.3.2. Cancers invasifs ou infiltrants.....	22
III.4. Le dépistage du cancer du sein.....	23
III.5. La mammographie.....	24
III.6. Les descripteurs de la mammographie.....	27
III.6.1. La forme.....	27
III.6.2. La texture.....	27
III.7. Conclusion.....	28
IV.1. Introduction.....	30
IV.2. Détection et classification assistées par ordinateur (CAdE / CAdx) des tumeurs dans la mammographie.....	30
IV.2.1. Détection des tumeurs assistée par ordinateur (CAdE).....	32
IV.2.2. Diagnostic des tumeurs assisté par ordinateur (CAdx).....	32
IV.3. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.....	32
IV.4. L'avantage d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CAdx)	34
IV.5. Conclusion.....	35
<i>Conclusion générale</i>	60
<i>Bibliographie</i>	62

Liste des figures

Fig I.1: Les techniques de l'analyse d'image.....	2
Fig I.2: Technique de calcul d'un local binary patterns.....	6
Fig I.3: Calcul de la matrice de co-occurrence ainsi que celle normalisée avec d = 1 et $\varnothing = 0$	7
Fig II.1: Forme d'une loi puissance dans un repère bi-logarithmique.....	11
Fig II.2: Codage de l'image Lena avec n=4, 6, 9, 12, 16.....	14
Fig II.3: Codage d'un motif 3x3 d'une image par la méthode des rangs généraux : en (a) nous exposons le motif original, et en (b) le motif codé par la méthode des rangs généraux.....	15
Fig II.4: image 12-B de la base MIAS et la courbe Zipf correspondant.....	16
Fig II.5: la courbe Zipf inverse de la même image au-dessus.....	17
Fig III.1: Sein touché par un cancer. Source.....	21
Fig III.2: Présentation d'un cancer du sein non invasif et un autre invasif.....	23
Fig III.3: Positionnement du sein pour la réalisation d'une mammographie.....	25
Fig III.4: Mammographie ne portant aucune anomalie.	25
Fig III.5: Mammographie portant une tumeur bénigne.....	26
Fig III.6: Mammographie portant une tumeur maligne.....	26
Fig III.7: Les différentes formes relatives à une masse : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière.....	27
Fig IV.1: Architecture d'un système d'indexation et de recherche des images par le contenu (CBIR).....	33
Fig V.1 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : N zone d'intérêt sans tumeur, B zone d'intérêt portant une tumeur bénigne, et M zone d'intérêt portant une tumeur maligne.....	39

Fig V.2 Obtention d'un LBP pour chaque pixel de la zone d'intérêt.....	42
Fig V.3 L'histogramme des LBP.....	43
Fig V.4 Pseudocode de l'algorithme génétique de base.....	44
Fig V.5 Illustration d'un croisement en un point qui sépare le génome de deux solutions à un rythme arbitraire pointez (ici au milieu) et réassemblez-les pour obtenir deux nouvelles solutions appelées enfants.....	45
Fig V.6 Cycle d'initialisation de l'algorithme génétique, crossover, mutation, fitness calcul, sélection et résiliation.....	46
Fig V.7 Sélection des descripteurs les plus discriminants par les algorithmes Génétiques.....	47
Fig V.8 Architecture global du système Mammo-ZipfLBP.....	51
Fig V.9 Fenêtre Login du système «Mammo-ZipfLBP».....	55
Fig V.10 Interface principale du système «Mammo-ZipfLBP».....	56
Fig V.11 Fenêtre des détails de réalisation de ce travail.....	57
Fig V.12 Fenêtre du processus de décision médicale issue de l'utilisation de la sortie de Mammo-ZipfLBP.....	58
Fig V.13 Processus de décision médicale issue de l'utilisation de la sortie de Mammo-ZipfLBP avec K=5 d'une requête maligne.....	59

Liste des tableaux

Tableau V.1 Variation de la précision de classification du système «Mammo-ZipfLBP» pour diverses valeurs des K plus proche voisins.....	53
Tableau V.2 Performance du système «Mammo-ZipfLBP» avec et sans l'utilisation des algorithmes génétiques.....	53
Tableau V.3 L'approche proposée vis-à-vis d'autres approches.....	54

Liste des acronymes

BI-RADS Breast Imaging-Reporting and Data System
CAdE Computer Aided Detection
CAdx Computer Aided Diagnosis
CBMIIR Content Based Mammogram Image Indexing and Retrieval
CBIR Content Based Image Retrieval
GA Algorithmes Génétiques
K-PPV k- plus proches voisins
LBP Local Binary Pattern
MATLAB Matrix Laboratory
MIAS Mammographic Image Analysis Society
ROI Region Of Interest
SVM Support Vector Machine
TF Transformée de Fourier
TFD Transformée de Fourier Discrète
TCD Transformée de Cosinus Discrète

Introduction générale

Au fil des dernières décennies, les femmes dans le monde entier, sont devenues la cible du cancer du sein consistant en l'une des maladies les plus fréquemment diagnostiqué et qui peut même leur causer la mort. En effet, les meilleures méthodes de prévention sont le diagnostic et le traitement précoce de cette maladie dangereuse afin d'améliorer le traitement et de diminuer la mortalité.

Donc, travailler sur le développement de méthodes fiables comme le diagnostic assisté par ordinateur était en ordre. En effet, des systèmes d'aide au diagnostic médical destinés aux radiologues ont été conçus l'aidant dans la prise de décision surtout dans des cas où les radiologues rapportent plusieurs interprétations pour la même image mammaire.

La problématique qui a été posée dans ce travail est l'analyse des images mammaires par des méthodes non linéaires qui sont adaptées à la structure complexe d'une image au lieu d'appliquer de simples méthodes linéaires qui posent des limites. Pour cela, nous nous intéressons aux lois puissance : Zipf et Zipf inverse pour l'analyse de la texture des images mammaires. En effet, nous proposons une fusion des lois de Zipf et de Zipf inverse avec l'approche LBP (Local Binary Patterns) dans l'objectif d'explorer l'apport complémentaire que peut générer ces deux approches de caractérisation de la texture afin d'essayer d'améliorer les résultats obtenus par l'application des lois de Zipf et de Zipf inverse individuellement.

Le but de notre travail est le développement d'un système d'indexation et de recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein. En effet, la différence principale entre un système d'aide au diagnostic médical du cancer du sein assisté par ordinateur et se basant sur la sortie d'un classifieur (CADx) et un système CBMIIR se situe dans la présentation du diagnostic. Nous suggérons d'améliorer la performance du diagnostic assisté par ordinateur (CADx) basé sur un résultat numérique offert par un classifieur et dont les radiologues ne sont pas accoutumées avec un diagnostic à base de cas traités et stockés préalablement dans une base de donnée.

Nous avons organisé notre mémoire en cinq chapitres :

Dans le chapitre I, nous introduirons les notions et concepts de base de l'analyse d'image et de la vision par ordinateur. Par la suite, nous détaillerons l'analyse de la texture dans les images et les méthodes utilisées dans ce sens.

Dans le chapitre II, nous aborderons les lois puissances en se concentrerons sur deux lois : Zipf et Zipf inverse en analyse d'image.

Au fil du chapitre III, nous introduirons le cancer du sein : définitions, types, causes, ainsi que l'outil de dépistage qui est la mammographie.

Dans le chapitre IV, nous présenterons les techniques assistées par ordinateur d'aide au diagnostic médical du cancer du sein ainsi que l'indexation et la recherche des mammographies par le contenu (CBMIIR) avec quelques travaux connexes réalisés dans ce contexte.

Durant le chapitre V, nous exposerons notre conception pour le développement du système d'indexation et de recherche des mammographies par le contenu et les résultats obtenus.

Nous terminerons ce travail par une conclusion générale résumant notre contribution tout en proposant les perspectives de notre travail.

Chapitre I

Analyse d'image et vision par ordinateur

I.1. Introduction

Durant les dernières années, la vision par ordinateur s'est révolutionnée dans le domaine d'analyse d'image en abordant des problèmes dont les humains résolvent à leur insu. En effet, l'enjeu principal est l'extraction de caractéristiques saillantes sous forme d'attributs aidant à la reconnaissance des objets présents dans une image.

Dans notre travail, nous nous intéressons au descripteur de la texture d'une image désignant un attribut visuel important dans la vision par ordinateur

Nous présentons dans le premier chapitre les notions de base de l'analyse d'image et de la vision par ordinateur. De plus, nous évoquons les différentes méthodes d'analyse de la texture. Effectivement, notre travail principal porte sur la caractérisation de la texture dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

I.2. Définition de l'analyse d'image

L'analyse d'images est un outil fondamental pour la reconnaissance d'objets. Les travaux pionniers de l'analyse d'image s'intéressaient principalement à des applications de détection de défauts, de segmentation ou de classification comme mentionné sur la [figI.1](#) où nous exposons les principales techniques d'analyse d'image en citant quelques approches de l'état de l'art pour chacune des techniques.[\[1\]](#).

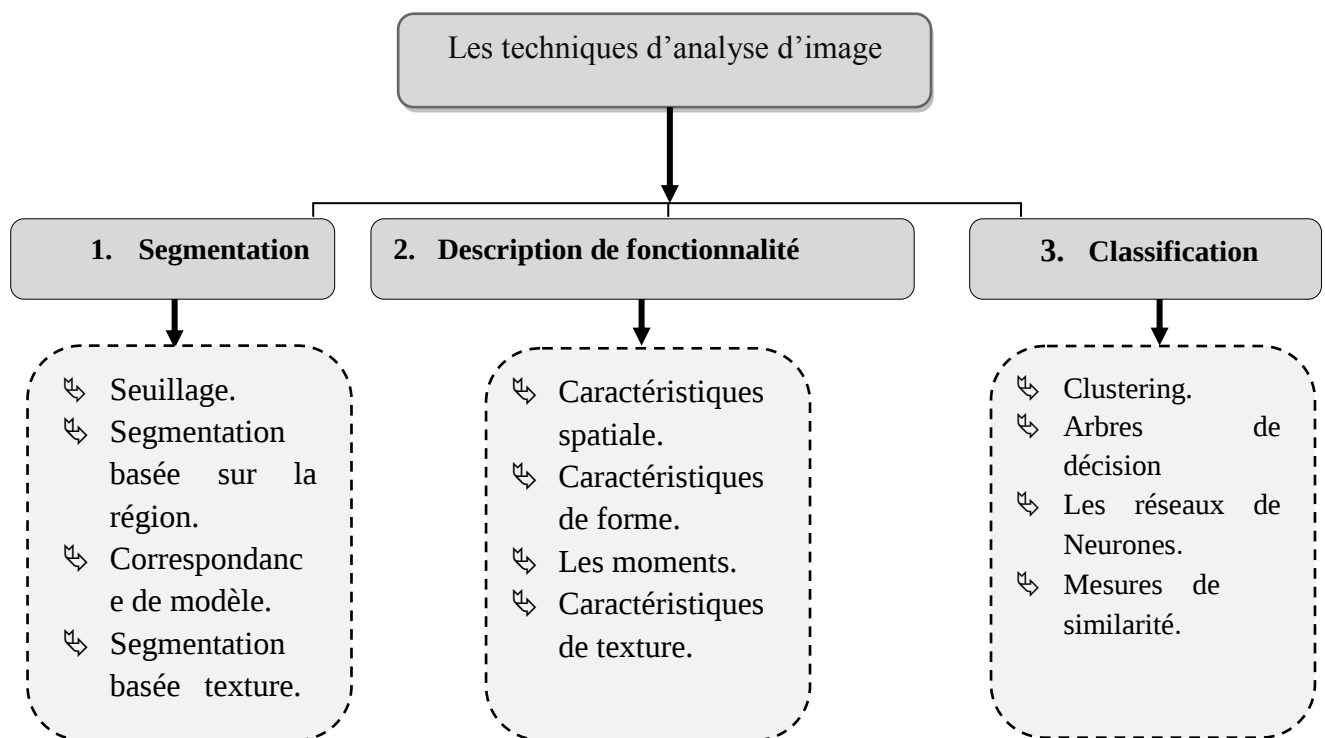


Fig I.1: Les techniques de l'analyse d'image [\[5\]](#).

En revanche, actuellement, l'analyse d'image devient de plus en plus importante et évidente ainsi qu'avec le moindre coût [2] du fait de son appuie sur la technologie de la vision artificielle [3]. En effet, la croissance explosive des plates-formes matérielles ainsi que les cadres logiciels ont engendrés des avancées surprenantes dans l'analyse des images numériques.

Notons que le processus d'analyse d'image renvoie généralement des valeurs numériques sur les caractéristiques de l'image étant évidemment adaptés à la classification. En effet, ce dernier génère une sortie sous forme d'un nombre ou une décision et non pas une image.

Les attributs sélectionnés doivent être suffisamment saillants pour la phase de reconnaissance des objets ainsi que leur localisation. Comprendre automatiquement le contenu d'une image est une problématique d'actualité dans la discipline de la vision artificielle.

I.2.1. Catégories de l'analyse d'images

L'analyse d'images regroupe plusieurs catégories [4] [6] citons:

a) Analyse de bas-niveaux d'image

Ces processus nécessitent très peu d'informations sur le contenu des images. Nous distinguons le processus de filtrage, celui d'amélioration et de restauration d'images. L'ultime but de ces processus est l'extraction d'indices saillants à partir des images.

b) Analyse d'intermédiaire niveau d'image

Le niveau intermédiaire est le niveau fonctionnel ayant comme rôle le passage d'une représentation purement basée sur l'image à une autre représentation d'objet. Ceci en identifiant et extrayant les régions présentant un potentiel intérêt pour un traitement ultérieur de haut niveau. Dans ce sens, nous faisons appels à des algorithmes de segmentation, ainsi que des algorithmes d'extraction de descripteurs [6].

c) Analyse de haut-niveaux d'image

Elle fonctionne en aval de celle de bas niveaux en nécessitant des informations sur le contenu des images. En effet, elle concerne la reconstruction tridimensionnelle, la reconnaissance de formes et les processus cognitifs.

d) Analyse de prise de décision d'image

C'est le niveau de prise de décision constituant le dernier niveau de traitement, ce niveau se base sur la récolte des informations acquises lors des étapes précédentes pour en générer des décisions de haut niveau [6].

I.3. Définition de la vision par ordinateur

L'enjeu majeur consiste en la tentation d'imiter la vision humaine au biais de composants électroniques. En effet, ceci consiste en le traitement des données visuelles par des modèles fondés sur la géométrie, la physique, la biologie, les statistiques ainsi que la théorie d'apprentissage.

La vision par ordinateur a fait l'objet de diverses application, citons l'analyse d'image médicales (détection de tumeur, analyse chromosomique) ou la robotique industrielle (reconnaissance et interprétation d'objets en scène) [6].

Un système d'analyse d'image de vision par ordinateur est structuré comme un pipeline de traitement de données ou un processus de traitement d'information, basé sur une gamme d'algorithmes regroupés en niveaux fonctionnels.

Ils existent des caractéristiques similaires dans les algorithmes de chaque niveau. Dipert et al. Identifient trois niveaux principaux dans, [7] à savoir:

- ↳ Acquisition d'image et optimisation dans le premier niveau.
- ↳ Convertir des pixels en objets dans le second.
- ↳ Analyse et reconnaissance des objets dans le dernier.

Un stimulus est présenté à l'entrée du système et l'information ainsi reçue est traitée afin d'en extraire une représentation symbolique. Ceci, présentera l'entrée pour un processus de traitement de plus haut niveau et l'information subira des traitements jusqu'à l'atteinte d'une compréhension raisonnable du monde environnant en permettant au système de vision d'agir comme par exemple de se diriger vers un but identifié [8].

I.4. Concept de l'analyse de la texture d'une image

La phase d'extraction de descripteurs d'image est essentielle dans de nombreuses applications relatives à l'analyse d'image. D'autre part, en vision par ordinateur, l'enjeu majeur réside dans l'extraction de descripteurs numériques discriminants. En effet, l'analyse de la texture présente un grand intérêt en vision humaine et en psychophysique, d'un autre

coté, les caractéristiques de texture peuvent ne pas être facilement perçues par la vision humaine.

I.4.1. Désignation du terme « Texture »

L'analyse de la texture consiste souvent à explorer des éléments de texture puis à dénombrer la répétitivité de ces éléments de base [6]. En outre, cette analyse fait appel à une classe des règles et des procédures mathématiques déterminant les variations spatiales au sein des images comme outil à fin de dégager l'information.

Donc, les textures nous fournissent des informations sur l'agencement spatial des intensités d'une image [12].

I.4.2. Les méthodes d'analyse de la texture

Suite à l'extraction de points d'intérêt à partir d'une image, l'étape succédant cette phase consiste en un codage dans un descripteur discriminant et insensible du contenu ou de la structure extraite. Notons que le descripteur doit être insensible aux changements qu'une image pourra subir comme la rotation par exemple.

Les méthodes d'analyse de la texture peuvent être classées en quatre groupes mentionnés ci-dessous:

I.4.2.1. Méthodes structurelles

Ces méthodes définissent fondamentalement les textures comme une répétition régulière d'un motif au sein d'une image [9]. En effet, le principe consiste à définir tout d'abord les éléments de texture (bords, polygones Voronoi) puis à calculer des descripteurs statistiques ou morphologiques par rapport à des lois de placement spécifiques des motifs.

Local Binary Pattern (LBP) est un descripteur de texture le plus célèbre dans les méthodes structurelles, le principe consiste à décrire la texture comme un histogramme de LBP. En effet, les modèles binaires représentent les relations d'intensité entre un pixel et ses voisins. Pour chaque pixel d'image, un LBP est obtenu en binarisant sa région voisine et l'intensité de pixel sera utilisée comme seuil, par la suite, le modèle binaire résultant est converti en un nombre décimal. A la fin, un histogramme est généré en tenant compte des occurrences de tous les LBP de l'image.

Nous présentons sur la [fig I.2](#) deux exemples de calcul d'un local binary patterns relatif à un pixel dans un voisinage 3x3.

Notons que les LBP présentent l'avantage de la caractérisation exacte et robuste de l'information locale, en effet, ils ont été utilisés avec brio pour l'identification ainsi que la classification des textures biologiques.

Le LBP classique et les descripteurs de forme ont été utilisés pour classer les cellules lymphocytes et réaliser le diagnostic de la leucémie lymphoblastique aigue sur des images de microscopie optique d'échantillons de sang [9].

D'autres variantes de la formulation de base de LPB ont été proposées la littérature mentionnées dans [9].

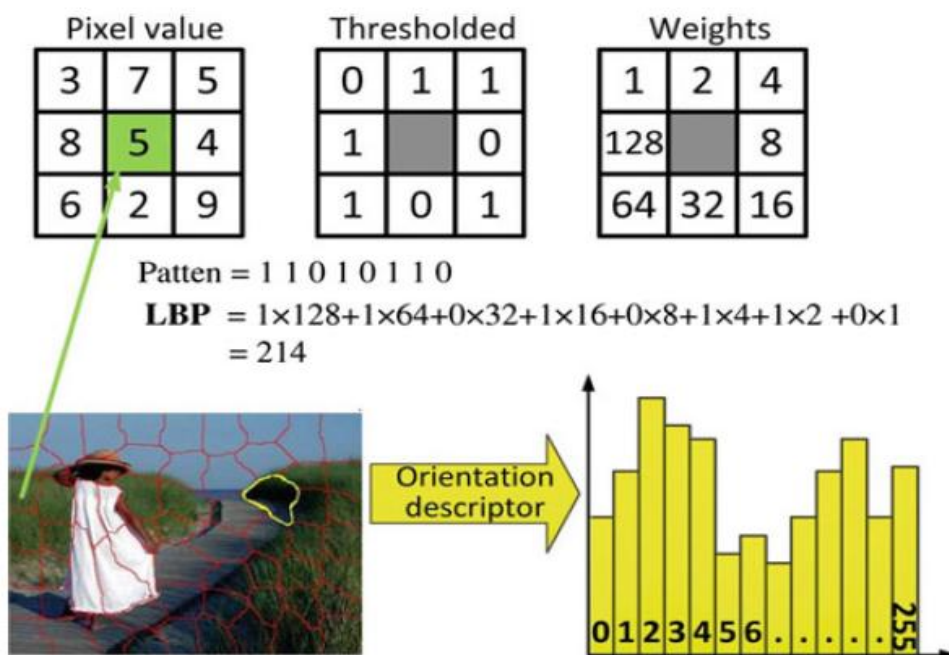
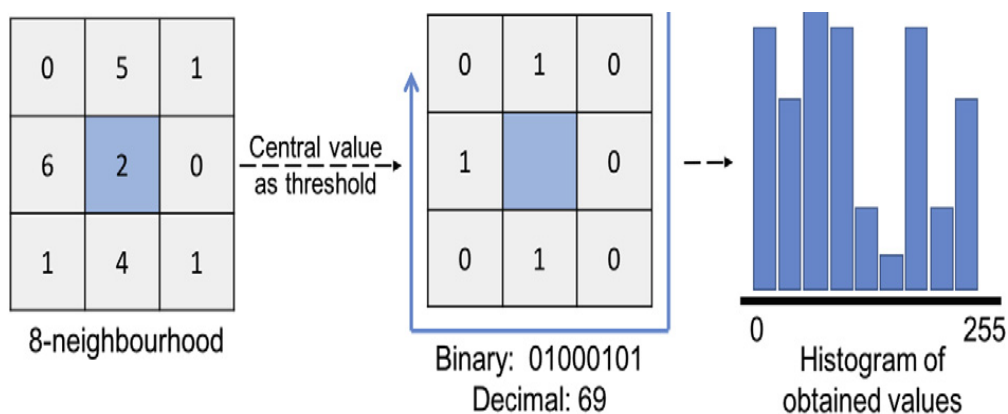


Fig I.2: Technique de calcul d'un local binary patterns. [9]

I.4.2.2. Méthodes statistiques

La texture est caractérisée comme une répétition déterministe des sous-modèles ainsi qu'une distribution spatiale non déterministe des d'intensité. Mathématiquement parlant, cela consiste en des statistiques de premier ou de deuxième ordre. En effet, les statistiques de premier ordre sont basées sur la ressemblance des pixels portant certaines valeurs d'intensité; néanmoins, ils ne sont pas assez puissantes pour reconnaître des textures complexes, citons l'exemple de l'histogramme des intensités d'une image où deux images peuvent avoir le même histogramme sans être identiques.

D'autre part, nous distinguons les méthodes statistiques du second ordre où la matrice de co-occurrence explore la ressemblance partagée de deux pixels aléatoires possédant les mêmes valeurs d'intensité au sein de l'image.

Nous présentons sur la [fig I.3](#) un exemple de calcul de la matrice de co-occurrence [9].

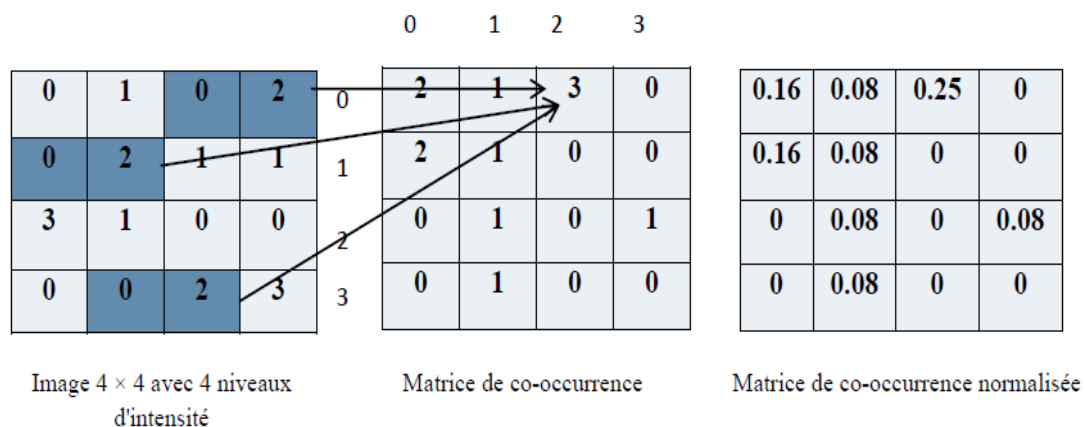


Fig I.3: Calcul de la matrice de co-occurrence ainsi que celle normalisée avec $d = 1$ et $\theta = 0$. [9]

I.4.2.3. Méthodes spectrales

Ces approches se sont inspirées du fait que le système visuel humain décrit des images dans le domaine fréquentiel [10]. L'ultime but consiste à convertir l'image en une nouvelle image en fonction des caractéristiques des fréquences spatiales de l'intensité des pixels [11]. Par la suite, les descripteurs de texture sont extraits à partir d'images filtrées, sur divers domaines.

Nous citons les plus utilisés:

- a) Filtres de domaine spatial :** Les méthodes issues du domaine spatial naïf dépendent sur des techniques simples de détection de contour (Sobel), puis l'extraction de la densité des bords dans l'image filtrée est effectuée et sera considérée comme un

descripteur de texture. Même si cette démarche distingue les motifs grossiers des motifs fins, elle présente des limites dans le fonctionnement des textures non systématiques [6].

b) Filtres de domaine de fréquence : L'analyse de fréquence peut être réalisée, soit par la Transformée de Fourier Discrète (TFD) bidimensionnelle, ou à travers la transformée de Cosinus Discrète (TCD). En effet, ces deux approches extraient les composantes de fréquence spatiale des images [6].

Gabor et les transformées en ondelettes. Contrairement à la TFD et la TCD, les ondelettes réalisent des décompositions spatiotemporelles où la base sinusoïdale est modulée avec des fonctions de fenêtres de formes différentes. Notons que l'utilisation d'une fenêtre de largeur limitée génère une meilleure localisation dans le domaine spatial si nous la comparons aux décompositions de Fourier traditionnelles, aboutissant aux meilleures capacités de discrimination [6].

Les filtres de Gabor mettent en évidence la résolution de l'image, à la fois, dans le champ spatial et celui de fréquence. Néanmoins, ils ne donnent pas de description complète concernant les textures grossières. En effet, l'énergie de ces textures se concentre dans les sous-images des fréquences les plus basses.

I.4.2.4. Méthodes basées sur un modèle

Les méthodes d'analyse de la texture à base de modèles utilisent la construction d'un modèle d'image qui permet par la suite de décrire une texture ainsi que d'en générer.

Les modèles les plus recommandés sont:

a) Modèles autorégressifs

Appliquant une interaction locale directe entre les pixels d'image. Le vecteur des poids représente le modèle des paramètres [9].

b) Champs aléatoires

La texture est considérée comme un échantillon fini d'un processus aléatoire bidimensionnel étant représenté par ses paramètres statistiques. Ceci se traduit par le fait que la ressemblance d'une cellule spécifique d'un treillis étant dans un état fourni (certaine intensité) est décidée en fonction de l'état des cellules adjacentes [9].

c) Fractales

Une fractale est une abstraction mathématique dans lequel une collection à différentes échelles affiche un motif répétitif identique à chaque échelle. Effectivement, les

paramètres fractals sont considérés comme des instruments qui estiment l'irrégularité ou l'hétérogénéité d'arrangements spatiaux [9].

I.5. Conclusion

L'avancement rapide de la technologie et de l'informatique a généré le challenge du développement des techniques très précises d'analyse et de traitement d'images représentant l'outil de base pour la vision par ordinateur. Dans ce sens, l'ultime but réside dans l'extraction des informations complètes, utiles et sémantiques à partir des images. Malgré toutes les recherches développées jusqu'à présent dans le domaine de l'analyse de la texture d'une image, la définition du terme texture demeure très difficile.

Notre problématique consiste en la proposition d'une fusion de l'approche local binary patterns d'analyse de la texture d'image avec celle des lois puissance Zipf et Zipf inverse pour la caractérisation de la texture des zones d'intérêt extraites à partir des mammographies dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur. En effet, nous tentons d'aboutir à une analyse accentuée de la texture en associant les lois puissance aux LBP qui est un puissant descripteur textural dont le principal avantage est l'invariance aux changements d'éclairage sur l'image.

Dans le chapitre suivant, nous exposons les modèles de loi puissance de type: Zipf et Zipf inverse, spécialement, leur application dans le domaine de l'imagerie.

Chapitre II

Les Lois puissance

II.1. Introduction

Les lois de puissance ont apparues dans les années 1897 sous la forme d'une répartition des revenus personnels provenant des enquêtes de Pareto en faisant l'objet d'application dans plusieurs domaines scientifiques.

Parmi les savants qui ont souligné une contribution importante le linguiste George Zipf en 1949, qui a contribué la loi de Zipf et la loi de Zipf inverse [25].

Dans ce chapitre, nous allons introduire les lois de puissance, précisément, les lois de Zipf et de Zipf inverse, puis nous évoquerons l'adaptation nécessaire pour leur extension à notre domaine de recherche, à savoir l'imagerie médicale et la vision par ordinateur.

II.2. Lois de puissance

Une loi puissance est une fonction associant deux quantités tout en étant de la forme suivante : $y = ax^{-b}$ avec a et b des constantes.

En général, les lois puissance sont représentées dans un repère bi-logarithme où la représentation graphique d'une distribution en loi puissance soit sous la forme d'une droite comme mentionné sur la Fig II.1 [25].

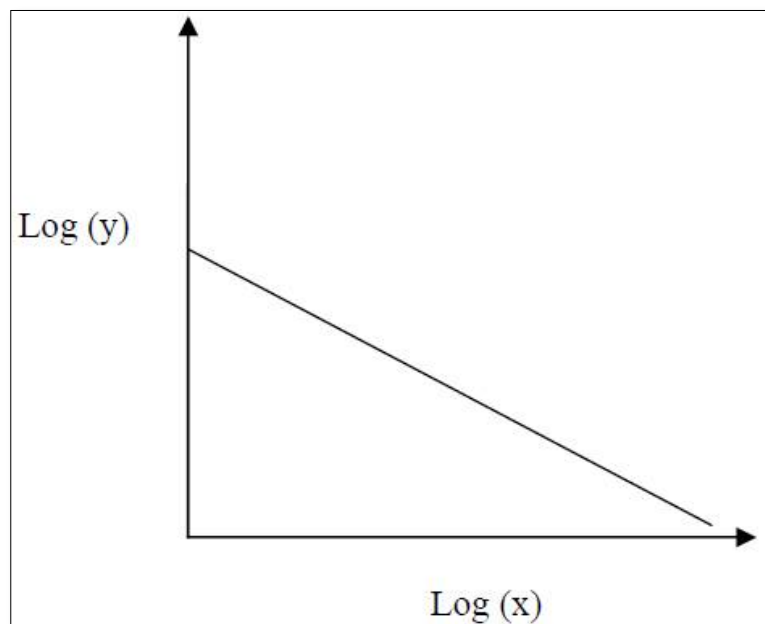


Fig II.1: Forme d'une loi puissance dans un repère bi-logarithmique.

II.3. Les principales de lois puissance

II.3.1. Loi de Pareto

Appelée également « loi des 80/20 », est l'une des premières lois puissance mise en place par l'économiste Wilfried Pareto en 1897.

Il extrait ses démonstrations à partir d'une définition simplifiée des variations de l'inégalité dans la distribution. En effet, suite à ses études sur la répartition des revenus personnel des individus ainsi que la richesse dans les principaux pays industrialisés, il a conclu que 80% des ressources appartenaient à seulement 20% de la population [25].

II.3.2. La loi de Zipf

En 1949, le linguiste George Zipf remarque quelque chose d'étrange par rapport à la fréquence à laquelle les gens utilisent les mots dans une langue. En effet, en classant les mots dans l'ordre décroissant de leur fréquence d'apparition, le mot classé numéro un a toujours été utilisé deux fois plus souvent que le mot de deuxième rang et trois fois plus souvent que le troisième rang. En effet, si nous désignons par F la fréquence d'apparition du mot de rang r dans la suite, nous exprimons cette relation par la formule suivante [25] :

$$F = a r^{-\alpha}$$

II.3.3. La loi de Zipf inverse

Pareil à la loi de Zipf, cette loi s'intéresse aux fréquences d'apparition des n -uplets d'un ensemble de symboles topologiquement ordonnés tandis que nous nous intéressons ici au nombre de n -uplets différents ayant la même fréquence d'apparition.

D'après la loi de Zipf inverse, le nombre I de n -uplets différents ayant une fréquence d'apparition f est donné par la formule suivante :

$$I(f) = a f^{-b}$$

❖ Dans cette formule, a et b sont des constantes positives [25].

II.4. Analyse d'images par les lois de Zipf et de Zipf inverse

L'extension de l'existence d'une répartition selon une loi puissance dans un cas bidimensionnel tel que l'image est très intéressante. En effet, nous désignons la modélisation de la fréquence d'apparition des motifs au sein de l'image selon des distributions en loi de puissance. Ceci nous mènera à caractériser la complexité structurelle des textures dans les images [25].

II.4.1. Analyse d'image par la loi de Zipf

Si nous considérons un motif $m \times m$ de l'image, la loi de Zipf décrit la distribution des fréquences d'apparition de chacun de ces motifs. Nous pouvons constater que si nous classons les fréquences $N_1, N_2, \dots, N_m$ d'apparition des motifs selon un ordre décroissant de leur fréquence d'apparition ($U_{\sigma(1)}, U_{\sigma(2)}, \dots, U_{\sigma(m)}$) et que nous associons à chaque motif son rang dans la séquence, les fréquences d'apparition des motifs sont données en fonction du rang par une loi puissance [13][30]:

$$N_{\sigma(i)} = k \cdot i^{-a}$$

❖ Dans cette formule, k et a sont des constantes.

II.4.2. Analyse d'image par la loi de Zipf inverse

A son tour, cette loi décrit les fréquences d'apparition des motifs de l'image topologiquement ordonnés, mais nous nous intéressons ici au nombre de motifs d'image différents ayant la même fréquence d'apparition.

D'après la loi de Zipf inverse, le nombre I de motifs différents ayant une fréquence d'apparition f est donné par la formule suivante [13][30]:

$$I(f) = a f^{-\gamma}$$

❖ Où a et γ sont des constantes

II.4.3. Codages nécessaires à l'analyse d'image par les lois de Zipf et de Zipf inverse

Le codage de l'image vise à réduire le nombre de motifs distincts et à augmenter la fréquence d'apparition d'un même motif au sein de l'image.

a) Codage des 9 Classes

Ce codage consiste en le rééchantillonnage de l'échelle des niveaux de gris de l'image. Pour cela, nous affectons une même valeur aux pixels ayant des niveaux de gris contenus dans un même intervalle de valeurs. En effet, nous partitionnons l'échelle [0, 255] des niveaux de gris en un nombre n de classes de largeurs égales et numérotés selon un ordre croissant de 0 à n .

Nous présentons sur la formule le principe de ce codage, en effet, nous attribuons à chaque pixel du motif la valeur $c(x, y)$ de sa classe en fonction de la valeur $g(x, y)$ de son niveau de gris, comme il est mentionné[13][30] :

$$c(x, y) = \text{int} \left[\frac{N * g(x, y)}{255} \right]$$

Notons que le choix du nombre de classes demeure difficile, en effet, nous devons choisir une valeur de n qui soit minime pour la réduction du nombre de motifs différents possibles, mais tout en étant suffisamment grande pour éviter l'altération de la structure de l'image.

Des essais ont été effectués pour diverses valeurs de n pour en extraire la valeur optimale satisfaisant au mieux les exigences contradictoires citées précédemment .La figure suivante montre une image codée avec des valeurs de n égales à 4, 6, 9, 12 et 16 :

$n=4$	$n=6$	$n=9$	$n=12$	$n=16$
				
2.621×10^5 motifs possibles	1.008×10^7 motifs possibles	3.874×10^8 motifs possibles	5.159×10^9 motifs possibles	6.872×10^{10} motifs possibles

Fig II.2: Codage de l'image Lena avec $n=4, 6, 9, 12, 16$. [31].

Les résultats expérimentaux ont montré que le choix d'une valeur de $n=9$ permet de préserver le contenu structurel de l'image tout en évitant une dégradation visible.

b) Codage des rangs généraux

Le codage respecte la vision d'une scène en se basant sur les variations locales de niveaux de gris, afin de mettre en évidence les différences relatives plutôt que les valeurs absolues.

Le principe est que les valeurs des niveaux de gris seront remplacées par leurs rangs lorsqu'elles sont triées en fonction des valeurs de niveau de gris en affectant la valeur 0 au niveau de gris le plus bas, et en incrémentant à chaque fois la valeur d'une unité. Il est utile de mentionner que nous affectons le même rang aux niveaux de gris ayant la même valeur.

Pour mieux cerner le principe de ce codage, La fig II.3 montre un exemple de motif codé avec les rangs généraux.

250	200	200	5	4	4
25	4	29	1	0	2
35	4	35	3	0	3
(a)			(b)		

Fig II.3: Codage d'un motif 3x3 d'une image par la méthode des rangs généraux : en (a) nous exposons le motif original, et en (b) le motif codé par la méthode des rangs généraux.

II.4.4. Construction des courbes de Zipf et de Zipf inverse

a) Construction de la courbe de Zipf

Nous procédons par un balayage séquentiel de l'image en considérant un masque de capture 3x3 du motif de l'image. Evidemment, le codage des motifs est appliqué en choisissant celui qui convient aux propriétés de l'image que nous cherchons à mettre en évidence.

L'étape à venir consiste à calculer le nombre d'occurrences de chacun des motifs distincts dans l'image, pour ceci, nous rangeons dans un tableau tous les motifs rencontrés tout en associant la fréquence d'apparition relative à chaque motif comme suit : en prenant en

considération chaque motif courant trouvé, nous le comparons aux autres déjà indexés dans le tableau : s'il existe déjà donc nous incrémentons sa fréquence d'apparition d'une unité, autrement, nous l'ajoutons au tableau en initialisant sa fréquence d'apparition à 1.

A la fin, nous trions les fréquences d'apparition des motifs selon un ordre décroissant et la fréquence de chaque motif est tracée en fonction de son rang dans un repère bi-logarithmique, où en abscisse le rang R des motifs et en ordonnée leur fréquence d'apparition. Nous présentons sur la fig II.4 la courbe de Zipf d'une image.[31]

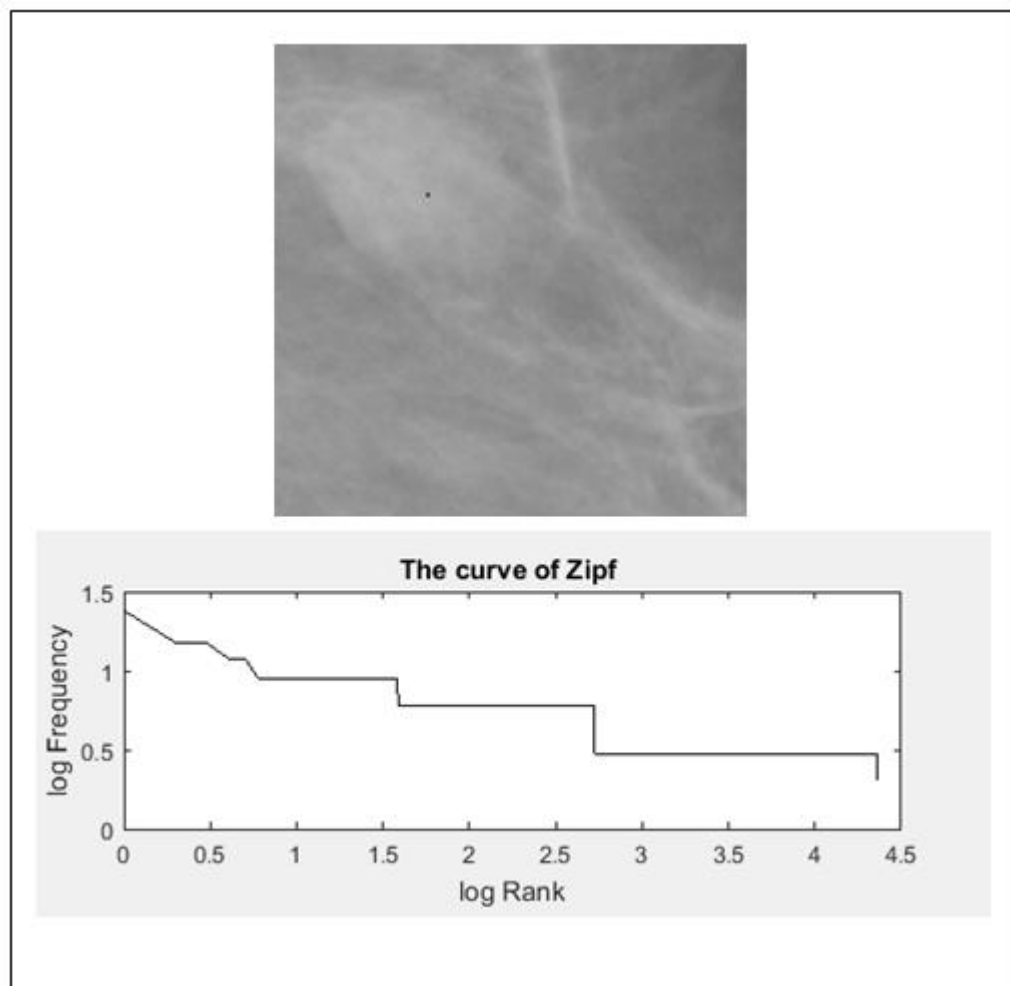


Fig II.4: image 12-B de la base MIAS et la courbe Zipf correspondant.

b) Construction de la courbe de Zipf inverse

Pour le traçage de la courbe de Zipf inverse d'une image, la première phase de l'algorithme est identique à celle utilisée pour le traçage de la courbe de Zipf. Effectivement, nous balayons l'image par un masque 3x3 tout en comparons le motif courant rencontré aux motifs déjà trouvés dans l'image, donc, soit le motif est présent et nous incrémentons sa fréquence d'apparition d'une unité, soit il n'est pas présent et nous le rajoutons dans le tableau.

La différence réside dans l'étape à venir, effectivement, nous dénombrons les motifs ayant la même fréquence d'apparition que la fréquence courante. Pour ceci, nous initialisons la fréquence cherchée à 1 et nous parcourons séquentiellement le tableau des motifs pour compter les motifs ayant la même fréquence que la fréquence courante.

Par la suite, l'algorithme est réitéré en incrémentant à chaque fois la fréquence d'une unité jusqu'à finir avec la fréquence maximale. [31]

Le nombre de motif est tracé en fonction de leur fréquence d'apparition dans un repère bi-logarithmique désignant la courbe de Zipf inverse.

Nous présentons sur la fig II.5 la courbe de Zipf inverse d'une image.

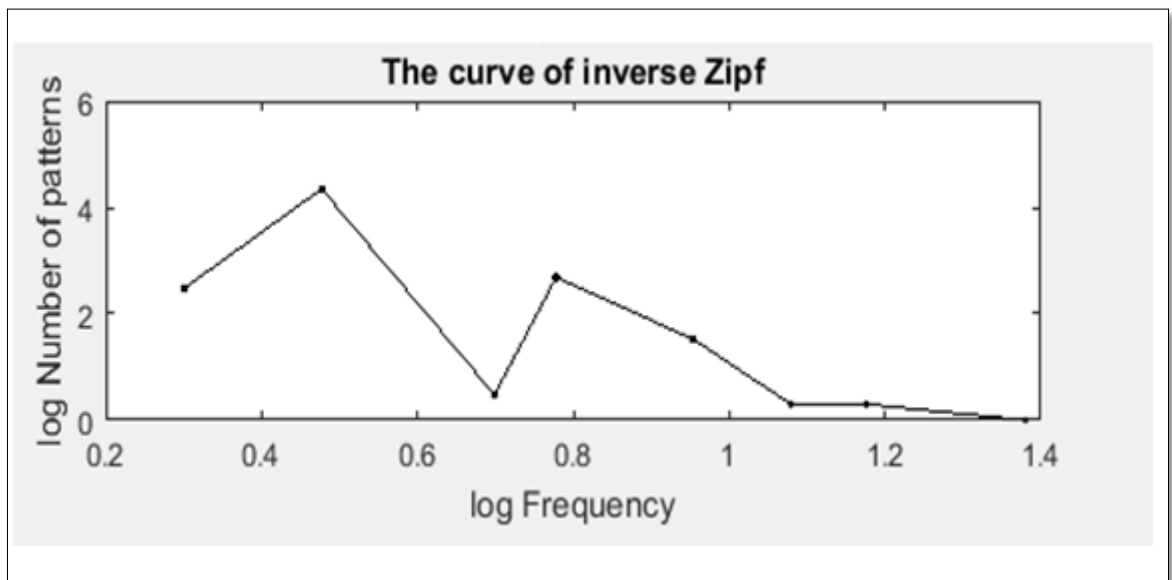


Fig II.5: la courbe Zipf inverse de la même image au-dessus

II.5. Conclusion

Les modèles de lois puissance comme la loi de Zipf et de Zipf inverse peuvent être appliqués à l'analyse d'image. En effet, nous définissons des motifs qui sont équivalents aux mots dans le cas de l'analyse de texte.

La pertinence des lois de Zipf et de Zipf inverses pour analyser la structure complexe que présente l'image numérique nous offre la possibilité de les appliquer au domaine de l'imagerie médicale pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur. A cet effet, le chapitre prochain évoquera des notions de bases sur le dépistage du cancer du sein.

Chapitre III

Cancer du sein : état de l'art

III.1. Introduction

Le cancer du sein est le premier cancer féminin en termes de fréquence. En effet, une femme sur dix sera atteinte de cette maladie au cours de sa vie et malheureusement une femme sur 25 décédera suite à un cancer du sein.

La détection précoce est fondamentale pour le traitement efficace du cancer du sein où nous distinguons la mammographie de dépistage comme l'outil le plus couramment utilisé par la communauté médicale augmentant les chances de survie jusqu'à 90% [6][16][17].

Le diagnostic assisté par ordinateur du cancer du sein devient une nécessité à cause de la croissance exponentielle du nombre de mammographies effectuées chaque année [18].

Dans ce chapitre, nous allons introduire le cancer du sein ainsi que les outils de son dépistage.

III.2. Le cancer du sein

Le cancer du sein est une maladie qui surgit de la multiplication de cellules mammaires anormales formant des masses. En 2015, 570 000 femmes sont décédées d'un cancer du sein.

Le diagnostic précoce du cancer du sein est essentiel à la survie du patient, dans ce sens, le dépistage par mammographie est le moyen le plus prometteur pour un diagnostic précoce [19]. Les cellules cancéreuses peuvent rester dans le sein ou se propager dans le corps à travers les vaisseaux sanguins ou lymphatiques. Généralement, le développement du cancer du sein peut prendre plusieurs mois voir même quelques années.

Le nombre de personnes atteintes a progressé légèrement au cours des 3 dernières décennies. Ceci dit, le taux de mortalité a continuellement diminué grâce aux progrès réalisés dans le dépistage, le diagnostic ainsi que le traitement. Mentionnons que les hommes peuvent aussi en être touchés; ils représentent 1 % de l'ensemble des cas [20].

Nous présentons sur la FigIII.1 un sein touché par le cancer où nous distinguons un carcinome lobulaire et un autre canalaire.

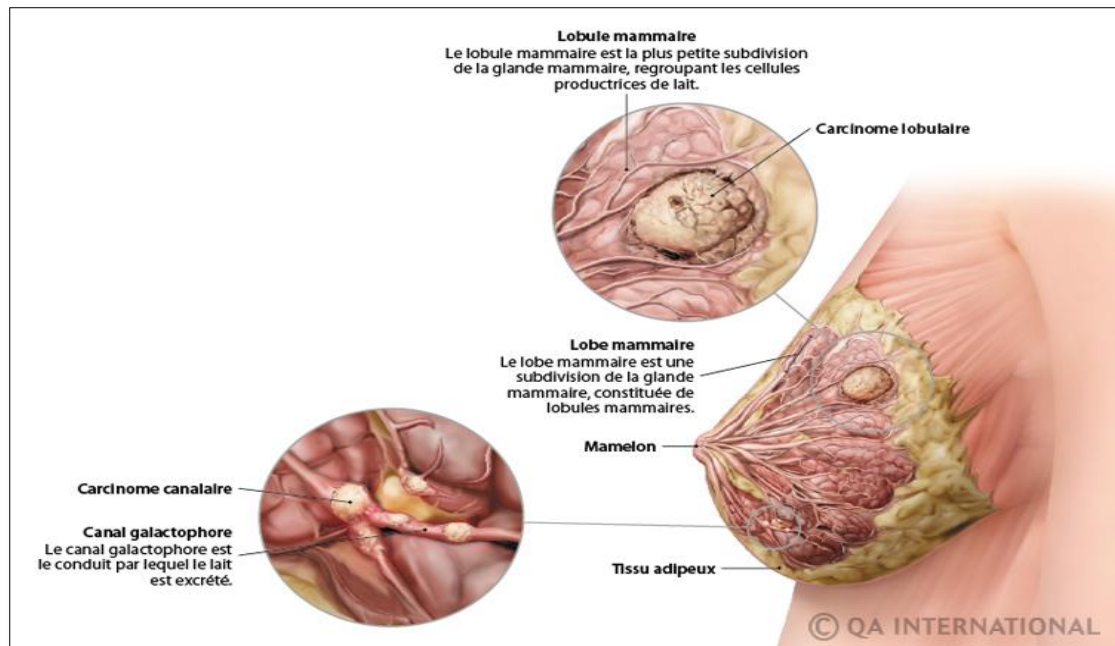


Fig III.1: Sein touché par un cancer. Source [21].

Il existe plusieurs facteurs de risque relatifs au cancer du sein. Néanmoins, dans la majorité des cas, il est impossible de connaître les raisons de son apparition dans le cas d'une personne en particulier. Nous allons citer quelques facteurs :

↳ L'âge

L'âge est le facteur de risque le plus répandu, en effet, 3 cancers du sein sur 4 apparaissent chez les femmes à partir de 50 ans et plus.

↳ Conditions hormonales

- ❖ Ne pas avoir eu d'enfant.
- ❖ Ayant le premier enfant après l'âge de 30 ans.
- ❖ Avoir eu ses premières règles avant 12 ans

↳ L'hérédité

Les causes héréditaires ou génétiques du cancer du sein concernent 5 à 10% des patientes. L'hérédité peut être la cause lorsque plusieurs parentes (la mère, la tante la sœur, un homme) sont ou bien ont été déjà atteints d'un cancer du sein, surtout s'ils ont été atteints avant la ménopause [23].

III.3. Les types du cancer du sein

Il existe deux principaux types de cancer du sein [20][22] :

III.3.1. Cancer non invasif

Le Carcinome canalaire in situ est le type le plus fréquent de cancer du sein non invasif chez la femme. En effet, il se forme à l'intérieur des canaux de lactation du sein. Nous diagnostiquons fréquemment ce genre de cancer depuis l'utilisation plus répandue de la mammographie

III.3.2. Cancers invasifs ou infiltrants

Ce type de cancer envahit les tissus avoisinant les canaux de lactation mais reste à l'intérieur du sein. Ceci dit, si la tumeur n'est pas traitée, elle pourra se disséminer vers d'autres parties du corps (l'os, les poumons ou le foie), générant des métastases. Nous distinguons :

a) Carcinome canalaire

Il se forme dans les canaux de lactation. Les cellules cancéreuses traversent la paroi des canaux.

b) Carcinome lobulaire

Les cellules cancéreuses apparaissent dans les lobules regroupés dans les lobes. Par la suite, elles traversent la paroi des lobules et se disséminent au-dedans des tissus environnants.

c) Carcinome inflammatoire

C'est un type rare se caractérisant principalement par un sein devenant rouge, enflé et chaud. Mentionnons que ce type de cancer progresse plus rapidement et est plus difficile à traiter.

d) Autres carcinomes

Médullaires, colloïdes ou muscineux, tubulaires, papillaires. Ces types de cancer du sein sont rares. Les différences principales entre ces types de cancer se basent sur le type de cellules touchées.

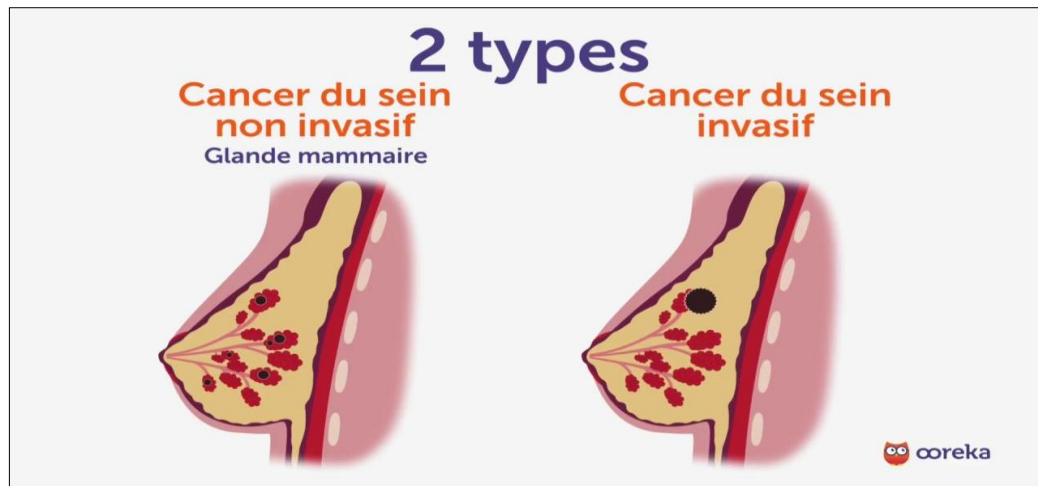


Fig III.2: Présentation d'un cancer du sein non invasif et un autre invasif. Source [21].

III.4. Le dépistage du cancer du sein

Le processus de dépistage aide à détecter le plus tôt possible le cancer dans le but de limiter sa progression, voir plus, de le guérir.

Nous cherchons toute aberration survenant au sein : bosse, écoulement, rétraction du mamelon, douleur, etc. Ceci au moyen de test clinique effectué aux seins par un professionnel du domaine de la santé pour la détection d'une masse.

L'ultime but du dépistage du cancer du sein est de baisser la mortalité en repérant une lésion avant l'apparition de symptômes, et donc de détecter des cancers non évolués. En effet, le cancer du sein détecté à un stade précoce nécessitera des traitements moins agressifs.

Il est utile de mentionner qu'à partir de 50 ans, le risque de l'atteinte du cancer du sein accroît nettement. A cet effet, dans le cas de progression d'un cancer du sein sans l'accomplissement régulier du dépistage, le cancer sera diagnostiqué à un stade développé. En effet, cela réduira les chances de guérison et fera subir la lourdeur des traitements, leur dangerosité parfois, et leurs séquelles aux patientes [22][24].

Si nous évoquons quelques symptômes du cancer du sein, nous commençons par le fait qu'il est tout à fait normal que l'aspect des seins change au fil des années, néanmoins, des modifications inhabituelles sont à observer. En effet, il est impératif de consulter un médecin dès la constatation des points suivants [22] :

- ↳ L'apparition d'une boule ou d'une grosseur dans le sein ou dans les aisselles.
- ↳ La modification de la peau : rétraction, rougeur, œdème ou aspect de peau d'orange.

- ↳ Le changement du mamelon ou de l'aréole qui est zone environnant le mamelon : rétraction, changement de coloration, suintement ou écoulement.
- ↳ Le changement de la forme des seins.

III.5. La mammographie

Le dispositif réalisant la mammographie examine tous les tissus mammaires via un tube à rayons X basse énergie et un système de compression mammaire. En effet, la mammographie est obtenue suite à une projection du volume du sein sur le plan relatif à l'image.

Les données existantes sur l'anatomie et la transparence à la radio, affirment que l'apparence globale de la mammographie est sombre tandis que les zones en présence de groupes ou composés de calcium sont nettement plus prononcées.

La mammographie est réalisée par un technicien en radiologie au moyen d'un appareil de radiologie consistant en la mammographie. En effet, les seins sont placés entre deux plaques qui se resserrent et les compriment pendant quelques secondes pour l'obtention de deux clichés par sein. Cette pression exercée par les deux plaques comprimant le sein est nécessaire pour garantir une bonne qualité des clichés en étalant le mieux possible le sein.

Par la suite, le radiologue interprète les clichés de la mammographie et pourra réaliser en complément davantage de clichés d'agrandissement dans le but d'affiner la lecture des images. Nous exposons sur la fig III.3 le positionnement du sein dans l'équipement de la réalisation d'une mammographie [18].

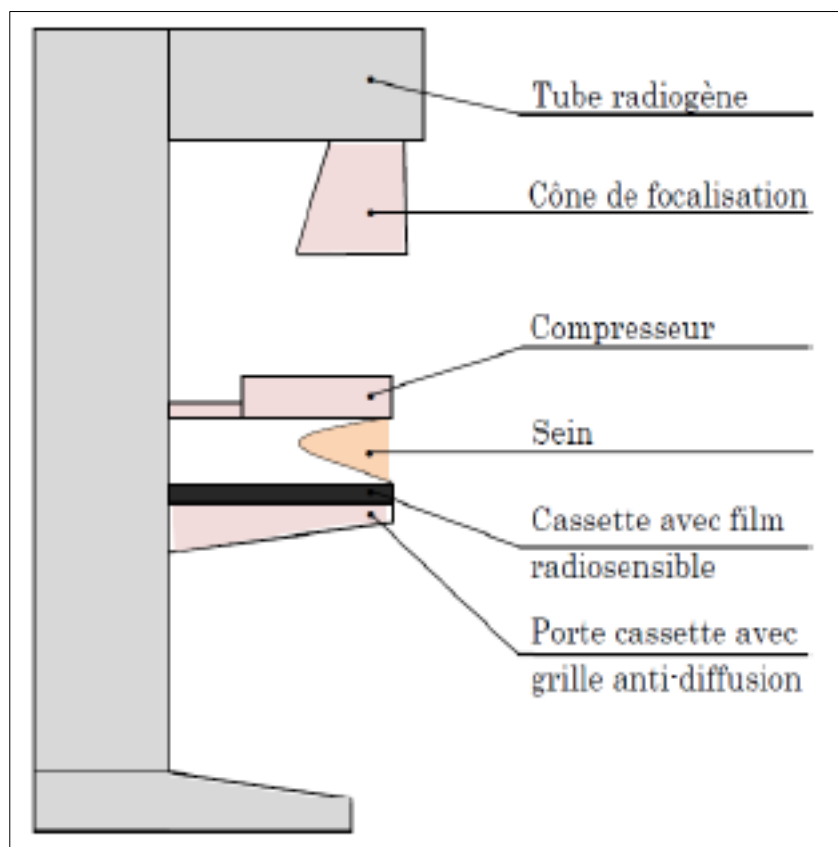


Fig III.3: Positionnement du sein pour la réalisation d'une mammographie [18].

Nous allons exposer sur la fig III.4, un exemple d'une mammographie de tissu sein ne possédant aucune anomalie.



Fig III.4: Mammographie ne portant aucune anomalie.

Tandis que sur la fig III.5, nous exposons une mammographie possédant une tumeur bénigne.

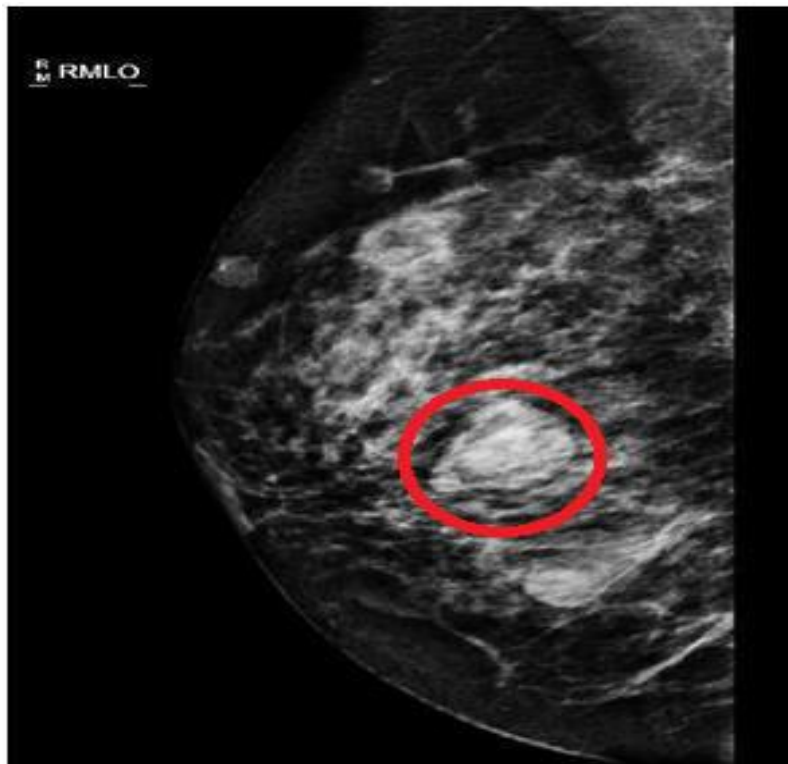


Fig III.5: Mammographie portant une tumeur bénigne.

Par ailleurs, nous exposons sur la fig III.6, une mammographie portant une tumeur maligne.



Fig III.6: Mammographie portant une tumeur maligne.

III.6. Les descripteurs de la mammographie

L'ultime but de la description d'image est d'extraire les caractéristiques qui décrivent au mieux quantitativement ou qualitativement les objets situés au sein de l'image. Ce processus est assuré en utilisant la couleur, la texture ou bien la forme.

Nous allons citer les différents types de descripteurs :

III.6.1. La forme

Selon la description du BIRADS[26] les masses mammaires peuvent présenter la forme ronde, ovale, lobulée ou irrégulière comme mentionné sur la fig III.7 [6] [18].

- ↪ **Ronde** : c'est une masse sphérique, circulaire ou bien globuleuse.
- ↪ **Ovale** : c'est une masse qui présente une forme elliptique comme la forme d'œuf.
- ↪ **Lobulée** : c'est une masse qui présente une légère ondulation.
- ↪ **Irrégulière** : c'est une masse dont la forme est aléatoire et ne peut en aucun cas être caractérisée par les termes précédents.

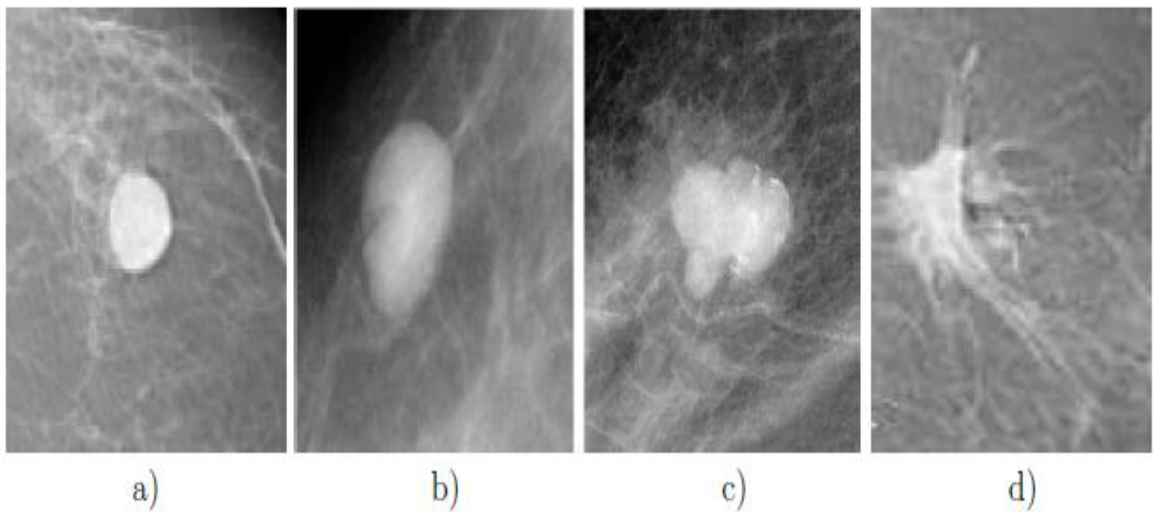


Fig III.7: Les différentes formes relatives à une masse : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière [18].

III.6.2. La texture

L'analyse de texture a prouvé sa pertinence pour la distinction entre sujets normaux et ceux pathologiques au cours de la classification de diverses conditions pathologiques.

Nous distinguons les annotations de texture suivante : fine ou grossière. Malgré que l'analyse de la texture soit d'une énorme importance vis-à-vis la vision physique humaine,

néanmoins, les caractéristiques de l'analyse de la texture ne sont pas facilement perçues et quantifiées par cette dernière.

III.7. Conclusion

Le cancer du sein est le type de cancer le plus courant chez les femmes dans le monde. Dans ce sens, l'outil le plus efficace pour la détection et le diagnostic demeure la mammographie. Néanmoins, l'interprétation de la mammographie présente une tâche répétitive et évidemment sujette aux erreurs.

Dans le prochain chapitre, nous allons discuter le principe du diagnostic médical du cancer du sein assisté par ordinateur pour offrir un deuxième avis aux médecins durant l'élaboration du diagnostic.

Chapitre IV

*Aide au diagnostic du cancer du sein
assisté par ordinateur*

IV.1. Introduction

Le diagnostic assisté par ordinateur, nommé également aide à la décision médicale par ordinateur, fait l'objet de nombreux travaux médicaux. L'ultime but est la progression des résultats obtenus lors des diagnostics médicaux effectués par imagerie, en effet, dû au faible contraste des mammographies ou bien à la fatigue visuelle qui touche les radiologues durant la lecture d'un énorme nombre de mammographies, les radiologues peuvent générer des erreurs.

Plusieurs systèmes de détection des tumeurs assistée par ordinateur (CADE) ainsi que de diagnostic des tumeurs assisté par ordinateur (CADx) ont été développés.

Dans ce chapitre, nous abordons les notions CADe et CADx ainsi que l'indexation et la recherche des mammographies par le contenu (CBMIIR) avec quelques travaux connexes réalisés dans ce contexte.

IV.2. Détection et classification assistées par ordinateur (CADE / CADx) des tumeurs dans la mammographie

Un système de détection ou de diagnostic assisté par ordinateur (CAD) est un système matériel et logiciel ayant comme objectif l'analyse des clichés médicaux aidant le spécialiste à la réalisation du diagnostic final. Concrètement, il s'agit d'un détecteur et d'un classifieur (KNN, SVM, arbres de décision, ...), en effet, le détecteur extrait les micro-calcifications et les masses, qu'elles soient de nature bénignes ou malignes. Tandis que le classifieur tri les zones d'intérêts en donnant une signification médicale à ce qui a été détecté.

Le CAD n'établit pas de diagnostic, Il donne une classification en catégories : bénin, malin, normal. Seul le médecin établira le diagnostic final en se basant sur l'aide offert par le système CAD désignant un outil pratique qui guide le spécialiste dans son travail. Ce système est classé en deux groupes: les systèmes de détection des tumeurs assistée par ordinateur (CADE) et les systèmes de diagnostic ou de classification des tumeurs assistés par ordinateur (CADx) [29].

Nous distinguons divers travaux connexes développant des systèmes de diagnostic assisté par ordinateur que ce soit pour la détection de masses (CADE) ou pour la classification (CADx).

Dans [32] les auteurs affirment que la texture est un outil précieux pour extraire des informations aidant au diagnostic en distinguant les tissus normaux des tissus pathologiques. Néanmoins, ils recommandent que la conception d'un système robuste d'aide au diagnostic

médical du cancer du sein (CAD) soit basée sur une large gamme de caractéristiques de texture ainsi que de fonctionnalités obtenues avec d'autres méthodes de traitement d'image.

Dans [33] les auteurs affirment que l'analyse manuelle d'image mammaire par un spécialiste est très compliquée du fait du faible contraste. Comme alternative, les auteurs proposent le diagnostic assisté par ordinateur du tissu mammaire basé sur l'analyse de la texture de toute la région de l'image, en partant des régions externes de l'image pour s'étendre aux régions intérieures. Pour cette tâche, ils utilisent des indices de diversité phylogénétique qui désigne une branche de la génétique traitant des modifications génétiques au sein des espèces animales ou végétales : la diversité pure ou la diversité phylogénétique; la somme des distances phylogénétiques; la distance moyenne du voisin le plus proche; la variabilité des espèces phylogénétiques; et la richesse en espèces phylogénétiques.

Dans [34, 35], les auteurs proposent des systèmes de CAD en mode B et en échographie moderne telles que l'élastographie et les ultrasons mammaires. Des caractéristiques de texture invariantes en niveaux de gris ont été utilisées où l'évaluation des tumeurs malignes était avec succès.

Une autre technique utilisée dans les systèmes de CAD dans le but d'améliorer la précision et l'efficacité des programmes de dépistage pour la détection et le diagnostic du cancer du sein est la fractale. L'approche fractale est basée sur l'identification de structures géométriques présentant une auto-similarité à différentes échelles [36, 37]. Dans [38], les auteurs utilisent la méthode fractale pour détecter et diagnostiquer les anomalies dans les mammographies. En effet, il est important que les tissus fibroglandulaires des deux seins soient symétriques et donc pour chaque densité asymétrique, une évaluation de la malignité devient nécessaire. Dans ce sens, les auteurs définissent de nouvelles caractéristiques fractales basées sur l'extraction d'informations asymétriques à partir de lésions où un classifieur SVM a été appliqué et une grande précision dans la classification des mammographies et des résultats de diagnostic a été obtenue.

Dans [39], les auteurs présentent un nouveau système de CAD qui détecte, segmente et classifie les masses mammaires à partir de mammographies. Plus spécifiquement pour la détection, ils ont proposé une cascade de méthodes d'apprentissage en profondeur pour sélectionner des hypothèses affinées en fonction de l'optimisation Bayésienne.

IV.2.1. Détection des tumeurs assistée par ordinateur (CAdE)

Les systèmes de détection des tumeurs assistée par ordinateur (CAdE) visent à localiser les lésions au sein des images médicales telles que les micro-assemblages agrégés et les masses. En effet, la première phase de ces systèmes consiste en la détection de régions suspectes. Les algorithmes les plus couramment appliqués dans ce sens pour l'extraction des zones d'intérêt (ROI: region of interest) se basent sur l'analyse des pixels.

Dans les techniques de détection basées sur les régions, les zones d'intérêt sont extraites suite à l'application de techniques de segmentation prenant en considération la morphologie ainsi que la taille des masses. Notons que leur complexité de calcul est nettement inférieure à celle des méthodes basées sur les pixels.

Par ailleurs, la seconde phase des algorithmes de détection de masse désigne la classification des zones d'intérêt (ROI) en tissu normal ou tissu pathologique [6, 13].

IV.2.2. Diagnostic des tumeurs assisté par ordinateur (CADx)

Les systèmes de diagnostic des tumeurs assisté par ordinateur (CADx) ont pour rôle de caractériser les lésions en réalisant la distinction entre les tumeurs bénignes et celle malignes. En effet, ceci permettra la réduction du nombre de recommandations biopsiques suggérés par les médecins concernant les tumeurs bénignes.

Suite à l'identification d'une tumeur par un radiologue ou un ordinateur (CAdE), le système de diagnostic assisté par ordinateur (CADx) utilise diverses approches de traitements d'images : l'extraction de caractéristiques des masses et des calcifications, la sélection des caractéristiques les plus discriminants et la classification pour la distinction entre les catégories des tissus mammaires [6].

IV.3. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur

Actuellement, les services de santé se focalisent fortement sur diverses données d'imagerie médicale qui ont augmenté considérablement en quantité vu l'augmentation rapide du nombre de visites médicales par jour. A cet effet, les chercheurs se sont beaucoup intéressés au domaine de développement de techniques structurées pour la gestion d'énormes

bases de données d'images médicales. En effet, l'enjeu principal est de rendre efficace : l'accès, la recherche et la récupération des images médicales.

Pour faire recours au problème du nombre énorme des images mammaires, l'approche d'indexation et de recherche d'image par le contenu semble la solution la plus adéquate par le biais de développement d'un système d'indexation et de recherche des mammographies par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval). Ces systèmes permettent l'accès rapide aux mammographies archivées, les plus similaires pathologiquement, à la mammographie en cours d'évaluation [13].

Les systèmes d'indexation et de recherche des mammographies par le contenu se basent sur le même principe des systèmes d'indexation et de recherche des images par le contenu (CBIR : Content Based Image Retrieval) où dans le module d'entrée, le vecteur de caractéristiques est extrait de l'image d'entrée qui sera stocké dans la base de données d'images. D'un autre côté, lorsqu'une image requête entre dans le module de requête, son vecteur de caractéristiques est extrait. Durant le processus de recherche, le vecteur de caractéristiques extrait est comparé aux vecteurs de caractéristiques stockés dans la base de données d'images. Par la suite, les images similaires sont récupérées et renvoyées par le processus de recherche en fonction de leur ressemblance à l'image requête suite au calcul d'une distance [6]. L'architecture du système d'indexation et de recherche des images par le contenu est présentée sur la Fig IV.1.

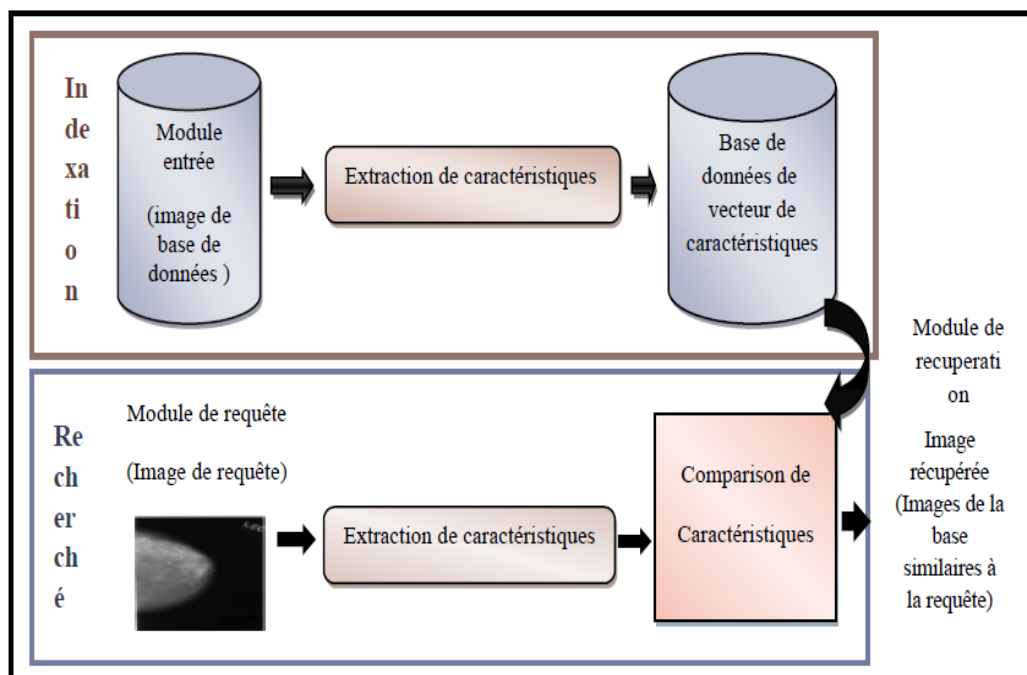


Fig IV.1: Architecture d'un système d'indexation et de recherche des images par le contenu (CBIR). Source [6].

IV.4. L'avantage d'un système d'indexation et de recherche de mammographies par le contenu (CBMIIR) par rapport à un système d'aide au diagnostic assisté par ordinateur (CADx)

Les systèmes d'aide au diagnostic assistés par ordinateur (CADx) offrent une aide substantielle aux radiologues durant la réalisation du processus de prise de décision. En effet, le but principal de tels systèmes est l'optimisation des efforts requis pour le diagnostic d'une tumeur. Si nous analysons la méthodologie principale de ces systèmes, nous distinguons l'extraction des caractéristiques à partir des lésions aidant dans la discrimination entre les masses bénignes et celles malignes ainsi que le tissu sein.

Cependant, les systèmes de CADx souffrent d'un taux plus élevé de faux positifs et de faux négatifs [13], ceci génère un nombre accru de biopsies inutiles causées par les performances relativement faibles des systèmes de CADe en matière de détection de masse.

Une autre tendance dans ce sens concerne les systèmes d'indexation et recherche des images mammaires par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval) offrant une décision basée sur les images mammaires récupérées durant le processus de recherche, étant les plus similaires à l'image mammaire requête. Nous affirmons que l'aide au diagnostic médical est visuelle dans le cas des CBMIIR ce qui augmente la sûreté des radiologues dans la prise de décision vu que le diagnostic est à base de cas similaires.

L'unique différence entre un système CADx et un système CBMIIR réside dans la présentation du diagnostic. En effet, l'enjeu principal consiste à renforcer la performance du diagnostic assisté par ordinateur (CADx) basé sur un résultat numérique donné par un classifieur et dont les radiologues ne sont pas familiarisés [13].

Dans les systèmes d'aide au diagnostic médical (CADx), un classifieur est optimisé sur un ensemble d'apprentissage, par la suite; la classification d'un nouveau cas d'entrée est réalisée. Ainsi, les radiologues obtiennent l'aide sous forme de symboles numériques indiquant par exemple la classe où appartient la région d'intérêt requête.

Par ailleurs, les radiologues se plaignent du fait que la sortie abstraite d'un classifieur n'est pas évidente pour eux du fait qu'ils se sentent plus confiant pendant la réalisation de leur diagnostic final lorsque des cas déjà diagnostiqués et similaires, au cas en cours de diagnostic, sont récupérés et affichés à partir de bases de données de référence ainsi que le résultat de la

classe à qui appartient le cas en cours d'analyse. En outre, notons la possibilité de l'indexation de nouveaux cas par les radiologues.

Il est utile de mentionner que les radiologues acquièrent leur formation en radiologie à travers des mammographies de lecture, en effet, leur expérience est construite par la lecture d'un énorme nombre de cas, qui les parviennent à reconnaître l'anatomie normale, l'anomalie bénigne ainsi que le tissu pathologique.

IV.5. Conclusion

L'exactitude des systèmes CAD pour l'aide au diagnostic médical du cancer du sein repose essentiellement sur la segmentation précise des lésions ainsi que l'extraction de descripteurs discriminants.

Le traitement des sujets atteints suite à l'utilisation des systèmes CAD, à un stade précoce du cancer du sein, augmente le taux de survie avec un rétablissement complet des patients. Evidemment ceci réduira le taux de mortalité généré suite à cette dangereuse maladie.

Chapitre V

Conception et Implémentation

V.1 Introduction

Durant la dernière décennie, les recherches dans le domaine médical au biais du traitement d'image et de la vision par ordinateur ont nettement évoluées. Dans ce sens, ce chapitre sera dédié au développement de notre système d'indexation et de recherche des mammographies par le contenu «Mammo-ZipfLBP» accomplissant la classification des tumeurs du sein par la technique du template-matching ainsi que la caractérisation de la texture des images mammaires par une fusion des lois de puissance : Zipf, Zipf inverse et des LBP.

V.2 Problématique

Notre problématique de recherche tente à explorer une technique efficace d'analyse de la texture des mammographies pour l'étendre vers la vision par ordinateur. Généralement les relations linéaires comme un polynôme ou une fraction rationnelle sont généralement les plus pratiqués pour l'analyse d'une image, néanmoins ces dernières sont restrictives vu qu'il est difficile de modéliser, avec exactitude, une structure compliquée telle que celle de l'image au moyen de simples modèles linéaires. Par ailleurs, des modèles plus complexes sont rarement employés, nous pouvons citer: les lois gaussiennes, les lois exponentielles et les lois puissance.

Dans [13] et [6], les auteurs ont contribué les lois de puissance Zipf et Zipf inverse pour la caractérisation de la texture des images mammaires dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur et ont pu démontrer la pertinence de ces lois. En effet, l'analyse de la texture au sein des mammographies par ces lois permet d'engendrer des descripteurs texturaux discriminants recueillant des particularités pour l'élaboration d'un diagnostic.

L'ultime but de notre travail consiste à travailler sur les perspectives de ces travaux en proposant une fusion des lois de Zipf et de Zipf inverse avec les LBP pour explorer l'apport complémentaire que peut générer ces deux approches dans le processus de la caractérisation de la texture.

Les images médicales jouent un rôle essentiel dans le diagnostic des patients, et quotidiennement, un nombre considérable des mammographies est réalisé. En effet, l'exploitation efficace de ces ressources demeure un défi pour les experts du domaine. Dans ce

sens, la nécessité de concevoir des systèmes précis qui permettent l'indexation et la recherche des images mammaires est d'une grande importance. Rappelons que les radiologues ne sont pas convaincus du résultat du diagnostic fourni à travers des symboles numériques issus des systèmes d'aide au diagnostic médical du cancer du sein se basant sur la sortie d'un classifieur (CADx). Donc, nous allons renforcer la performance du diagnostic assisté par ordinateur basé sur la sortie d'un classifieur (CADx) par l'alternative des systèmes d'indexation et de recherche des mammographies par le contenu (CBMIIR), où les radiologues se sentent plus confiants dans leur diagnostic élaboré à base de cas déjà diagnostiqués et similaires, au cas en cours d'analyse.

Nous allons développer un système d'indexation et de recherche des images mammaires basé sur la fusion des lois de Zipf et Zipf inverse avec les LBP pour la caractérisation de la texture des zones d'intérêt extraites à partir des mammographies pour effectuer un diagnostic assisté par ordinateur se basant sur la technique du template-matching.

V.3 Principe de la méthode basée fusion des lois de puissance : Zipf, Zipf inverse et les LBP pour la classification des zones d'intérêt des mammographies

Dans notre méthode, nous allons fusionner les descripteurs texturaux obtenus suite à l'analyse d'image par les lois de Zipf et de Zipf inverse avec ceux des LBP.

V.3.1 Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse

Les lois de Zipf et de Zipf inverse sont fondées sur une approche statistique et effectuent la quantification de la structure sous-jacente du contenu de l'image pour l'analyse ainsi que la caractérisation de la complexité structurelle de la texture d'une image.

Dans le but d'extraction des descripteurs de la texture au sein d'une image mammaire, les zones d'intérêts ont été segmentées selon une approche proposée dans [14]. Par la suite, un pré-traitement est appliqué aux pixels des zones d'intérêt en utilisant le codage des rangs généraux (mentionné dans la sous-section II.4.3) pour le traçage des courbes de Zipf et Zipf inverse, comme exposé sur la fig V.1 où un exemple de courbes de Zipf et Zipf inverse obtenu à partir de trois zones d'intérêts : l'une sans lésion, une autre portant une tumeur

bénigne et la dernière portant une tumeur maligne. Il est utile d'illustrer que dans le cas des images mammaires seines nous avons utilisé un découpage d'une partie du tissu sein.

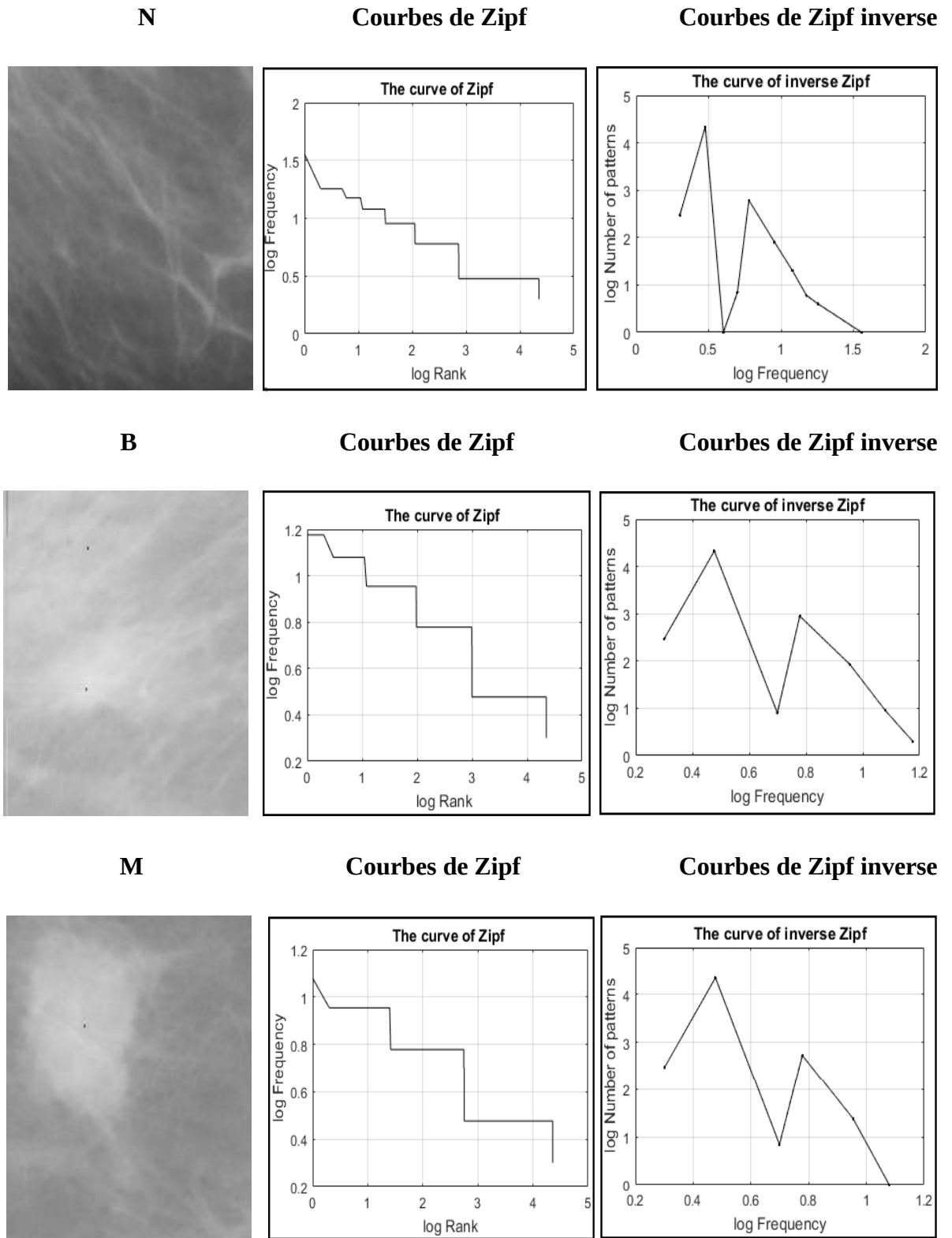


Fig V.1 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : **N** zone d'intérêt sans tumeur, **B** zone d'intérêt portant une tumeur bénigne, et **M** zone d'intérêt portant une tumeur maligne. Source [13]

- **Analyse des courbes de Zipf**

Nous observons une ordonnée à l'origine très haute du fait que la texture est homogène dans le cas d'une zone d'intérêt présentant un tissu normal du sein et nous distinguons donc un motif homogène se répétant plusieurs fois.

Alors que pour le cas où la zone d'intérêt présente une tumeur bénigne, nous observons une ordonnée à l'origine plus basse. L'ordonnée à l'origine de la courbe de Zipf pour la zone d'intérêt présentant une tumeur maligne est clairement plus basse que celles de la tumeur bénigne étant donné que la tumeur maligne présente une texture hétérogène causée par la nature invasive des tumeurs malignes, à ce fait, nous ne distinguons pas de sur-représentation du motif homogène, au contraire, plusieurs motifs hétérogènes provoquent que l'ordonnée à l'origine soit basse.

- **Analyse des courbes de Zipf inverse relatives**

Nous pouvons affirmer que la distinction entre les courbes de Zipf inverse est évidente pour la zone d'intérêt présentant un tissu normal, la zone d'intérêt présentant une tumeur bénigne et la zone d'intérêt présentant une tumeur maligne, où l'ordonnée à l'origine de la courbe de Zipf inverse de la tumeur bénigne est plus basse que celle de la courbe de Zipf inverse de la tumeur maligne par le fait que la région d'intérêt de la tumeur maligne a une texture complexe, donc, le nombre de motifs ayant une faible fréquence d'apparition est très élevé faisant que la courbe de Zipf inverse a un grand nombre de motifs qui n'apparaissent qu'une seule fois.

Les descripteurs que nous extrayons à partir des courbes de Zipf et de Zipf inverse sont les suivants :

- **Les pentes des courbes de Zipf et de Zipf inverse :**

La pente moyenne d'une courbe est le coefficient directeur de la droite des moindres carrés.

Elle est donnée par la formule V.1.

$$p = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (V.1)$$

- **L'aire délimitée par la courbe de Zipf :**

Nous calculons l'aire délimitée par la courbe de Zipf à partir des courbes de Zipf obtenues suite au codage de l'image par le codage des rangs généraux. Soit n le nombre de motifs de la courbe, f_i la fréquence et r_i le rang du motif i , l'aire de la courbe est donnée par la formule V.2.

$$A = \sum_{i=1}^{n-1} \frac{(f_i + f_{i+1})(r_{i+1} - r_i)}{2} \quad (V.2)$$

- **Entropie1 de la courbe de Zipf**

L'entropie relative aux motifs des images mammographies est définie par la formule:

$$H_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (V.3)$$

Dans cette formule, $f(r)$ représente la fréquence du motif pour la ligne r , T représente le nombre total de motifs différents, et nous utilisons un logarithme avec la base R .

- **Entropie2 de la courbe de Zipf**

L'entropie relative à la fréquence d'apparition des motifs est définie par la formule suivante:

$$H_f = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (V.4)$$

Dans cette formule, $I(f)$ représente le nombre de motifs distincts ayant une fréquence d'apparition égale à f et F représente le nombre entier d'occurrences des motifs dans l'image.

- **Les ordonnées à l'origine des courbes de Zipf et Zipf inverse**

- **Le constant alpha de la courbe de Zipf**

La loi de Zipf est fortement exprimée de la façon suivante: Quel que soit un motif appartenant à une image, la fréquence d'apparition de ce motif * son rang dans une liste ordonnée décroissante des fréquences d'apparition des motifs = constante.

V.3.2 Analyse et caractérisation de la texture des zones d'intérêt par les LBP

Pour chaque pixel d'image, un LBP est obtenu en binarisant sa région voisine et l'intensité de pixel sera utilisée comme seuil comme mentionné sur la fig V.2.

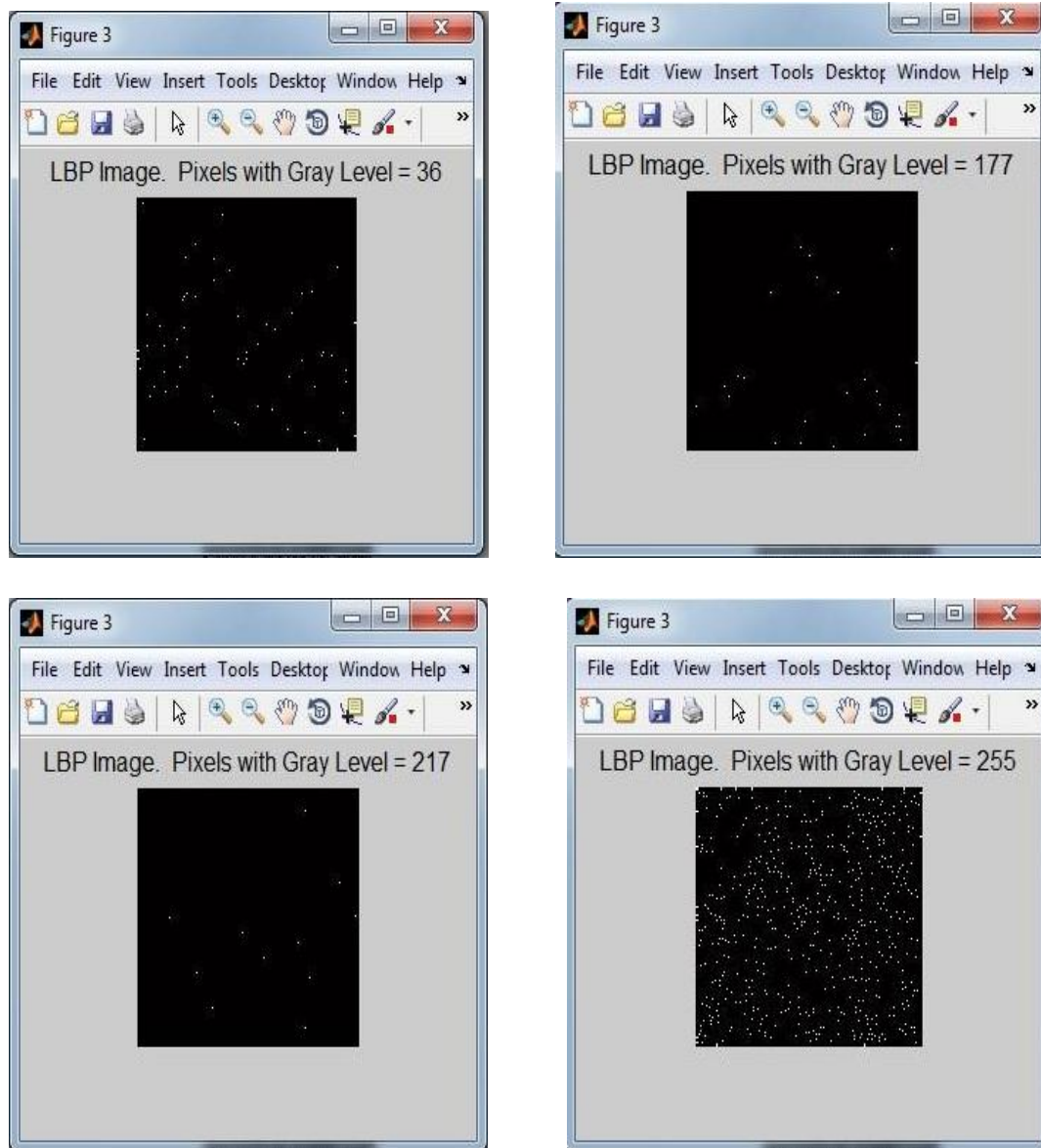


Fig V.2 Obtention d'un LBP pour chaque pixel de la zone d'intérêt

Par la suite, le modèle binaire résultant est converti en un nombre décimal. A la fin, un histogramme est généré en tenant compte des occurrences de tous les LPP de l'image comme mentionné sur la fig V.3, en effet, les descripteurs seront calculés à partir de cet histogramme.

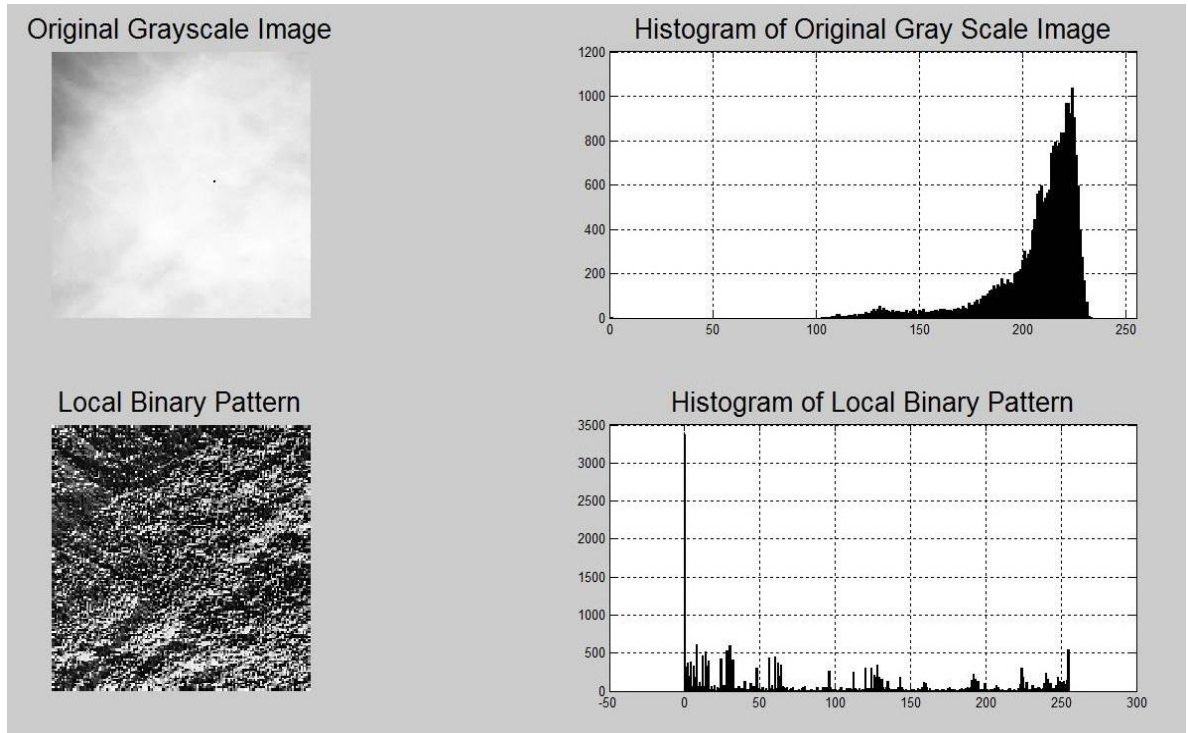


Fig V.3 L'histogramme des LBP.

V.4 Réduction de la dimension du vecteur descripteur par les algorithmes génétiques

V.4.1 Définition des algorithmes génétiques

Les algorithmes génétiques (GA) [27, 28] sont des approches de recherche heuristiques appartiennent à la famille des algorithmes évolutionnistes, ils sont des méthodes d'optimisation inspirées par la biologie. Au cours des dernières décennies, ils sont devenus des moyens exceptionnels pour résoudre les problèmes d'optimisation.

Les algorithmes génétiques sont la traduction du concept biologique de l'évolution en recettes algorithmiques. Ils utilisent la notion de sélection naturelle et l'appliquent à une population de solutions potentielles au problème donné. Ils appartiennent au domaine de l'informatique lié aux machines et aux programmes informatiques.

V.4.2 Algorithme génétique de base

L'algorithme génétique classique comme mentionné sur la fig V.4 est basé sur un ensemble de solutions candidates qui représentent une solution au problème d'optimisation que nous voulons résoudre. Ils étant basés sur des phénomènes biologiques, il convient de rappeler au préalable quelques termes de génétique. Se concentrant sur les concepts les plus importants tels que les populations, le schéma générationnel, le croisement (crossover), mutation, sélection. Sa représentation joue un rôle important, car la représentation détermine le choix des opérateurs génétiques. Les représentations sont généralement des listes de valeurs et sont plus généralement basées sur des ensembles de symboles. S'ils sont continus, on les appelle vecteurs, s'ils sont constitués de bits, ils s'appellent des chaînes de bits (bit string).

Au début, un ensemble de solutions, appelé population, est initialisé. Cette initialisation est recommandée pour couvrir de manière aléatoire tout l'espace de la solution ou pour modéliser et incorporer des connaissances expertes. La représentation détermine le processus d'initialisation. Pour la chaîne de bits (bit string) représentations, une combinaison aléatoire de zéros et de uns est raisonnable, par exemple le chromosome aléatoire initial **1001001001** en tant que chaîne de bits typique de longueur **10**.

Algorithm 1 Basic GENETIC ALGORITHM

```
1: initialize population
2: repeat
3:   repeat
4:     crossover
5:     mutation
6:     phenotype mapping
7:     fitness computation
8:   until population complete
9:   selection of parental population
10: until termination condition
```

Fig V.4 Pseudocode de l'algorithme génétique de base [27].

- **Crossover**

Crossover est un opérateur qui permet la combinaison du matériel génétique de deux solutions ou plus. Les opérateurs de crossover dans les algorithmes génétiques mettent en œuvre un mécanisme qui mélange le matériel génétique des parents.

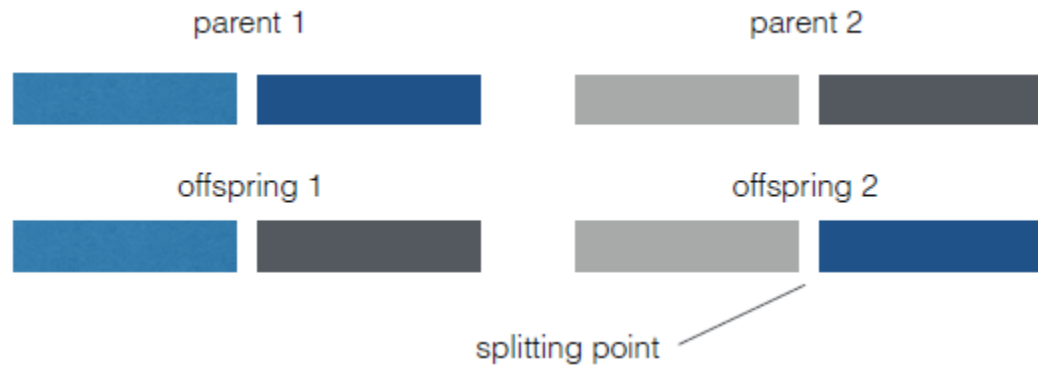


Fig V.5 Illustration d'un croisement en un point qui sépare le génome de deux solutions à un rythme arbitraire pointez (ici au milieu) et réassemblez-les pour obtenir deux nouvelles solutions appelées enfants [27].

- **Mutation**

Le deuxième protagoniste des algorithmes génétiques est la mutation. Les opérateurs de mutations changent une solution en les perturbant. La mutation est basée sur des changements aléatoires. La force de cette perturbation est appelée taux de mutation. Dans les espaces de solution continus, le taux de mutation est également appelé taille de pas.

- **Cartographie génotype-phénotype (Genotype-Phenotype Mapping)**

Après crossover et mutation, la nouvelle population de progéniture (les enfants) doit être évaluée. Chaque solution candidate doit être évaluée en fonction de sa capacité à résoudre le problème d'optimisation. La cartographie génotype-phénotype n'est pas toujours requise. Par exemple, dans l'optimisation continue, le génotype est la solution elle-même. Mais de nombreux autres processus de modélisation évolutifs nécessitent ce mappage.

- **Fitness**

Dans l'étape de calcul de la condition physique, le phénotype d'une solution est évalué sur une fonction de Fitness. La fonction fitness mesure la qualité des solutions générées par

l'algorithme génétique. La conception de la fonction fitness fait partie de la modélisation processus de toute l'approche d'optimisation. La performance d'un algorithme génétique dans la résolution d'un problème est généralement mesurée en termes du nombre d'évaluations requises de la fonction de fitness jusqu'à ce que l'optimum soit trouvé ou approximatif avec la précision souhaitée.

- **Sélection**

Pour permettre la convergence vers des solutions optimales, les meilleures solutions de progéniture doivent être sélectionnées pour être les parents de la nouvelle population parentale. Un excédent de solutions de progéniture est généré et les meilleurs sont sélectionnés pour progresser vers l'optimum. Ce processus de sélection est basé sur les valeurs de fitness de la population.

- **Termination**

La condition de terminaison définit la fin de la boucle évolutive principale. Souvent, l'algorithme génétique s'exécute sur un nombre prédéfini de générations. Le temps et le coût des évaluations de la fonction fitness peuvent limiter la durée du processus d'optimisation. Nous exposons sur la fig V.6 le cycle d'initialisation de l'algorithme génétique, crossover, mutation, fitness calcul, sélection et résiliation.

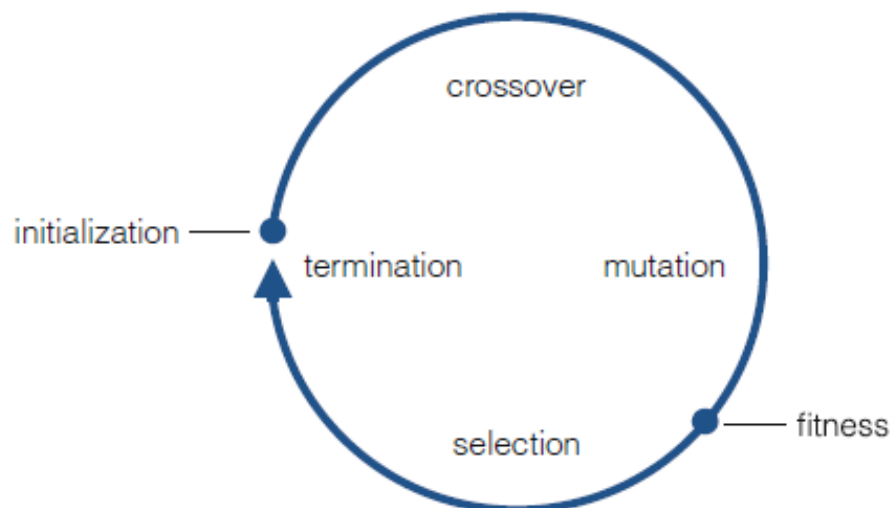


Fig V.6 Cycle d'initialisation de l'algorithme génétique, crossover, mutation, fitness calcul, sélection et résiliation [27].

Dans notre problème nous avons utilisé un algorithme génétique non déterministe dont le résultat (le nombre de descripteurs) est inconnu alors notre algorithme se base sur la corrélation des colonnes pour extraire les meilleurs descripteurs (les caractéristiques pertinentes) donc comme a été expliqué dans les sections précédentes et aussi les principes de l'algorithme sont tous appliqué comme se suit:

Nous avons utilisé trois classes la classe **AG** contient les paramètres de l'algorithme:

La taille de population (PopulationSize) = 10.

Nombre de générations (Generations) = 10.

Le type de population (PopulationType) = chaîne de bits (bitstring).

La classe Fit_fct contient la fonction de fitness dont nous avons présenté une fonction qui sélectionne les colonnes le plus corrélées et détermine les nouveaux descripteurs après l'appel de chromosome non vide et en utilisant les indices de colonnes sélectionnées et puis la fitness est calculée.

La classe **testAG** calcule le score de les matrices corrélées et les trie par les minimum corrélations pour extraire la meilleure matrice et puis prend la première moins corrélée.

Ensuite, nous obtenons une matrice réduite avec les descripteurs les plus discriminants. La Fig V.7 montre l'exécution de l'algorithme avec 10 générations et la sélection des meilleur 11 individu parmi 21.

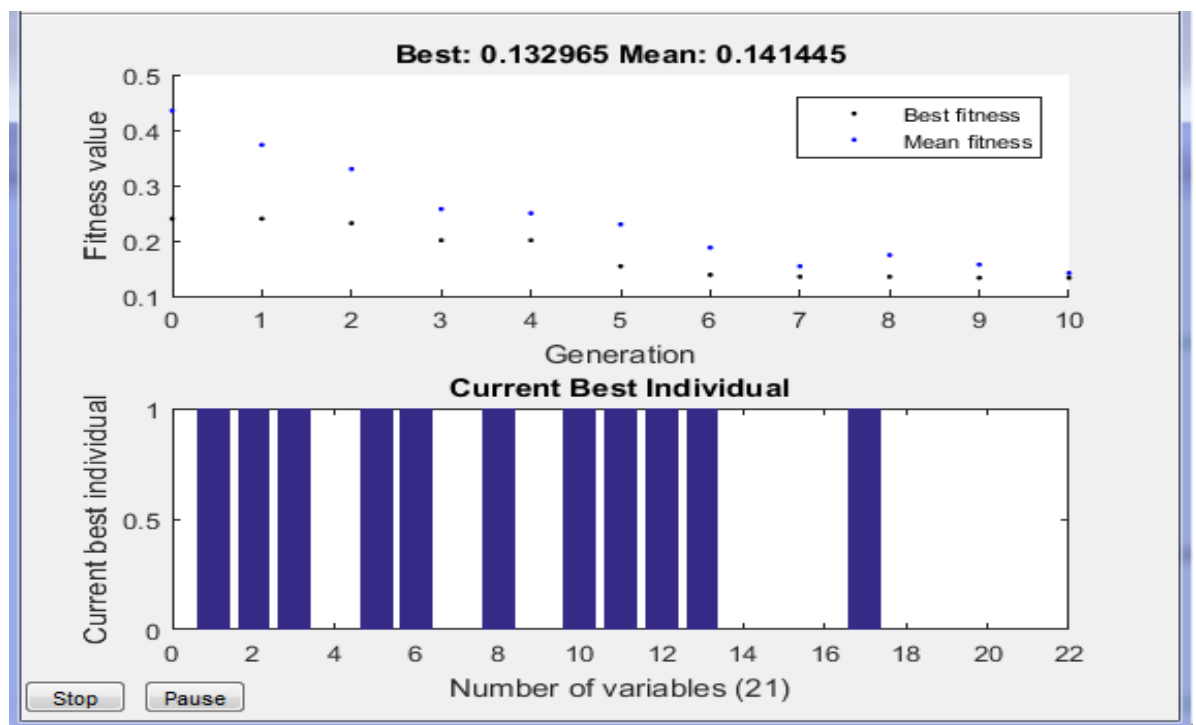


Fig V.7 Sélection des descripteurs les plus discriminants par les algorithmes génétiques.

V.5 Classification des zones d'intérêt

Nous allons utiliser simultanément, l'algorithme des KPPV ainsi que les machines à vecteur de support pour le processus de classification.

V.5.1 Classification basée Template-matching par l'algorithme des K plus proches voisins

Template-matching est une technique de traitement d'image numérique permettant de rechercher de petites parties d'une image qui correspondent à une image de modèle, donc cette étape vise à trier tous les descripteurs des zones d'intérêt indexées par rapport à la similarité avec le vecteur descripteur de la requête. Ceci est acquis en calculant la distance euclidienne entre deux vecteurs caractéristiques, sur laquelle se base l'algorithme des K plus proches voisins [6].

En abrégé k-NN ou KNN, de l'anglais k nearest neighbors [6, 15], est une méthode supervisée de classification, le classifieur KNN est basé sur l'idée d'estimer la classe d'une instance inconnue à partir de ses voisins. Il tente de regrouper les occurrences du vecteur de caractéristiques en classes disjointes en partant du principe que les occurrences de vecteurs de caractéristiques proches les uns des autres dans un espace de fonctions représentent des instances appartenant à la même classe. L'algorithme KNN calcule les distances entre le point et les points de l'ensemble de données d'apprentissage. La distance euclidienne est utilisée généralement comme mesure de similarité (ou fonction de distance), par la suite, il assigne le point à la classe majoritaire des K plus proches voisins. Les performances de classification du classifieur KNN dépendent de la valeur de K (où K est un entier).

Un point crucial de cet algorithme est la fonction de distance utilisée pour mesurer la proximité des objets. Il n'existe pas de distance/similarité universellement optimale et une bonne connaissance du problème traité guide généralement au choix de cette distance/similarité. Pour mesurer la distance entre les points A et B dans un espace de caractéristiques, différentes distances ont été utilisées dans la littérature, dans laquelle la fonction de distance euclidienne est le plus largement utilisé. Soit A et B représentés par des vecteurs de caractéristiques $A = (x_1, x_2, \dots, x_m)$ et $B = (y_1, y_2, \dots, y_m)$, où m est la dimensionnalité de l'espace caractéristique.

Pour calculer la distance entre A et B, la métrique euclidienne normalisée est définie comme suit :

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^m (x_i - y_i)^2}{m}} \quad (V.5)$$

V.5.2 Les machines à vecteurs de support (SVM)

Les machines à vecteurs de support [6, 18] est une méthode de classification binaire par apprentissage supervisé. Le principe de base des SVM consiste de ramener le problème de la discrimination à celui, linéaire, de la recherche d'un hyperplan optimal. Pour séparer les points appartenant à différentes classes, l'hyperplan optimal est celui qui a la distance maximale des points de données. Elle est basée sur l'utilisation de fonctions dites noyaux (kernel) qui permettent une séparation optimale des données.

Une caractéristique importante des fonctions du noyau et de l'espace noyau est que le produit interne entre deux vecteurs dans un espace de dimension supérieure peut être calculé sans projeter explicitement les vecteurs dans l'espace de dimension supérieure.

Pour deux classes données, le but du SVM est de trouver un classifieur qui va séparer les données et maximiser la distance entre ces deux classes, ce classifieur est appelé hyperplan. Il existe une multitude d'hyperplans séparant les deux classes, mais on retient l'hyperplan dont la distance minimale aux exemples d'apprentissage est maximale, on appelle cette distance marge. Les points les plus rapprochés utilisés pour la détermination de l'hyperplan sont appelés les vecteurs de support.

Il existe deux cas de modèles SVM: les cas linéairement séparables et les cas non-linéairement séparables. Les premiers sont les plus simples, car ils permettent de trouver facilement le classifieur linéaire.

Dans les cas non-linéairement séparables, un système de marges élastiques est alors mis en place pour tolérer des erreurs de classification et donc tolérer des vecteurs supports à l'intérieur de la marge.

V.6 L'architecture globale du système Mammo-ZipfLBP

La fig V.8 présente l'architecture de notre système Mammo-ZipfLBP. Nous approfondissons les étapes de notre système suggéré comme suit :

- **L'étape hors ligne**

- ✓ L'étude des zones d'intérêt par les lois de Zipf et de Zipf inverse afin de tracer les courbes de Zipf et Zipf inverse et ressortir les 8 descripteurs texturaux générés par ces lois.
- ✓ Extraire les 13 descripteurs texturaux générés par les LBP.
- ✓ Réduire la dimension du vecteur descripteur contenant à la fois les descripteurs des lois de Zipf et de Zipf inverse ainsi que ceux d'Haralick par les algorithmes génétiques.
- ✓ Indexation des vecteurs descripteurs dans base de données.

- **L'étape en ligne**

- ✓ Le radiologue introduit une zone d'intérêt requête, segmentée à partir d'une mammographie au système pour obtenir l'aide au diagnostic.
- ✓ L'analyse de la zone d'intérêt requête par les lois de Zipf et de Zipf et l'extraction des descripteurs texturaux.
- ✓ Obtentions des descripteurs texturaux de cette zone d'intérêt à partir des LBP.
- ✓ Application des algorithmes génétiques sur le vecteur descripteur global de la zone d'intérêt requête.
- ✓ Calcul de similarité (matching) entre le vecteur descripteur de la zone d'intérêt requête et les vecteurs descripteurs de toutes les zones d'intérêt dans la base de données par la distance Euclidienne intégrée dans l'algorithme des K plus proches voisin.
- ✓ Les cas renvoyés par le système suite à la recherche des K plus proches voisin de la région d'intérêt requête seront, désormais, prédicteurs de la classe pathologique de la requête en cours de diagnostic. En effet, nous attribuons à la zone d'intérêt requête la classe pathologique majoritairement renvoyée par le processus de recherche.
- ✓ Nous avons utilisé aussi le classificateur (SVM) pour le processus de classification.

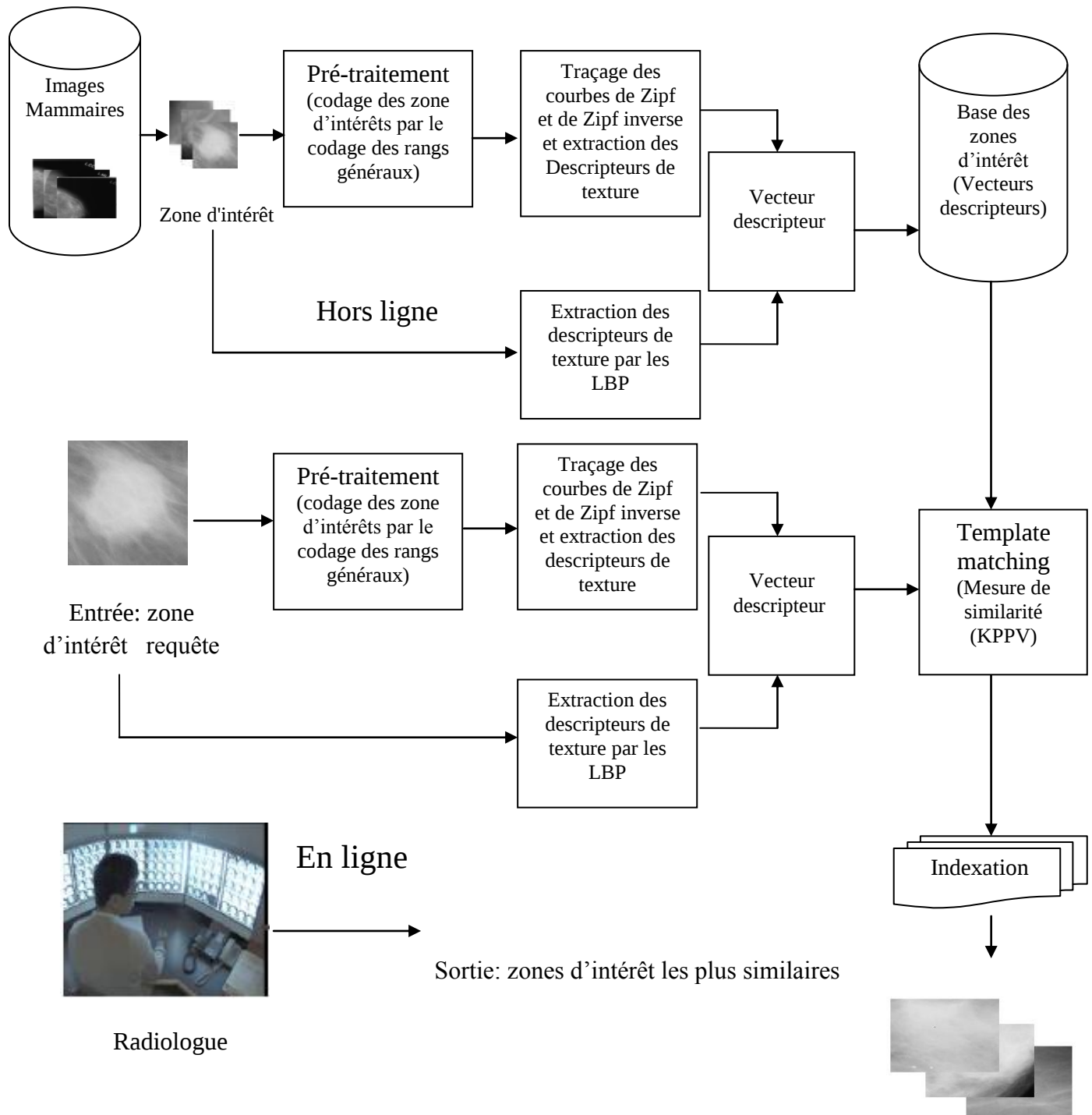


Fig V.8 Architecture globale du système Mammo-ZipfLBP.

V. 7 Implémentation

V.7.1 L'environnement de développement

Notre application a été réalisée sur un PC de type Dell Inspiron 15 4.00 GB de RAM sous Windows 7 Pro.

Les langages utilisés sont : Matlab version 2015a.

MATLAB (matrix laboratory) est un environnement informatique numérique multiparadigme. Langage de programmation propriétaire développé par MathWorks, MATLAB permet la manipulation de matrices, le traçage de fonctions et de données, l'implémentation d'algorithmes, la création d'interfaces utilisateur et l'interfaçage avec des programmes écrits en C, C ++, C #, Java, et Python.

On a choisi Matlab pour les raisons suivantes :

- Matlab facilite le traitement des images médicales de format de mammographie.
- Facilité la manipulation et d'assimilation syntaxe le Matlab.
- Il offre un niveau supérieur de précision mathématique pour les calculs complexes, et des fonctions mathématiques prédéfinies.
- Fournir une haute de stockage précision à l'information de l'image, l'arithmétique, ce qui rend les données d'image avec une plus grande précision flottante.

V.7.2 La base de données MIAS

L'approche proposée dans ce travail a été validée sur des images mammaires de la base MIAS (Mammograms Image Analysis Society). La base de données contient 120 images mammaires qui sont déjà diagnostiquées, où divisé en pathologie connues 59 images mammaire sans lésions, 37 images mammaires portant une tumeur bénignes et 24 portant une tumeur maligne. Notons que la méthode de segmentation des zones d'intérêt a été présentée dans [13].

V.7.3 Evaluation des performances

Pour l'évaluation des performances de notre système, nous avons divisé la base de données MIAS selon le principe de la validation croisée, qui est principalement utilisé dans les contextes où l'objectif est la prédiction lorsque nous désirer évaluer la performance d'un

modèle prédictif dans la pratique. Le principe consiste en la division de l'échantillon de taille n en deux sous échantillons, le premier dit d'apprentissage (supérieur à 60 % de l'échantillon) et le second dit de test.

À partir de ceci, nous avons divisé notre base de données MIAS en deux parties, la première ($> 60\%$ de l'échantillon) consistant en l'ensemble des données d'apprentissage (75 images mammaires) et nous avons consacré le reste de l'échantillon pour le test.

Nous avons d'abord calculé la précision de classification du système «Mammo-ZipfLBP» pour diverses valeurs des K plus proche voisins comme mentionné sur le tableau suivant :

Tableau V.1 Variation de la précision de classification du système «Mammo-ZipfLBP» pour diverses valeurs des K plus proche voisins.

Les valeurs de K	Précision de la classification (%)
2	64%
3	67%
5	69%
8	67%
9	68%

Nous pouvons affirmer que la précision de la classification varie entre 64% et 69% et que la meilleure performance du système proposé a été obtenue pour $k=5$.

Le tableau V.2 évoque les résultats de la précision de classification et le temps requis pour le système «Mammo-ZipfLBP» en utilisant les K -PPV pour $K=5$ et les SVM, avec et sans l'utilisation des algorithmes génétiques.

Tableau V.2 Performance du système «Mammo-ZipfLBP» avec et sans l'utilisation des algorithmes génétiques.

Les méthodes Les Paramètres	Application des algorithmes génétiques		Sans l'application des algorithmes génétiques	
	K-PPV ($K=5$)	SVM	K-PPV ($K=5$)	SVM
Précision de classification(%)	69%	76%	60%	70%
Temps requis	0.98281	0.5928	0.96721	3.8844

Les résultats montrent que la performance du système «Mammo-ZipfLBP» est meilleure lorsque nous appliquons les algorithmes génétiques pour l'obtention des descripteurs discriminants. En outre, le classifieur des SVM aboutit à une meilleure précision de classification (76%) par rapport au classifieur K-PPV (69%).

Nous avons comparé nos résultats obtenus avec d'autres approches comme indiqué sur le tableau suivant :

Tableau V.3 L'approche proposée vis-à-vis d'autres approches.

Approche	Technique	Base de données	Précision de classification(%)
Approche proposée : Zipf, Zipf inverse et les LBP	K-PPV	MIAS	69%
	SVM		76%
Zipf, Zipf inverse et les Ondelettes de Haar	K-PPV	MIAS	80%
	SVM		70%
Zipf, Zipf inverse et Haralick	K-PPV	MIAS	80%
	SVM		68%
Zipf, Zipf inverse et les SFTA	K-PPV	MIAS	62.5%
	SVM		60%

V.8 Présentation de quelques vues du système «Mammo-ZipfLBP»

Le système «Mammo-ZipfLBP» est doté d'une fenêtre login permettant d'accéder aux différentes fonctionnalités du système. Ceci est mentionné sur la Fig V.9.



Fig V.9 Fenêtre Login du système «Mammo-ZipfLBP»

En exécutant le bouton «Entrer», nous accédons à l'interface principale du système «Mammo-ZipfLBP» après avoir introduit le nom d'utilisateur et le mot de passe bien évidemment comme mentionné sur la Fig V.10.



Fig V.10 Interface principale du système «Mammo-ZipfLBP».

En exécutant le bouton «À propos», nous aurons des détails sur le travail proposé comme exposé sur la Fig V.11.



Fig V.11 Fenêtre des détails de réalisation de ce travail.

En exécutant le bouton «Entrer» de la fenêtre principale, la fenêtre de la fig V.12 apparaît.

En cliquant sur le bouton « Parcourir », le radiologue choisira la zone d'intérêt requête qu'il désire diagnostiquer comme mentionné sur la même figure.

Sur fig V.13, nous démontrons également la sortie du système « Mammo-ZipfLBP » proposé. Le radiologue a bien introduit la zone d'intérêt requête au système et cherche les régions d'intérêt similaires afin de fournir le résultat de la classification par les deux algorithmes des KPPV ainsi que les SVM : tumeur maligne ou bien bénigne ou tissu de parenchyme sein.

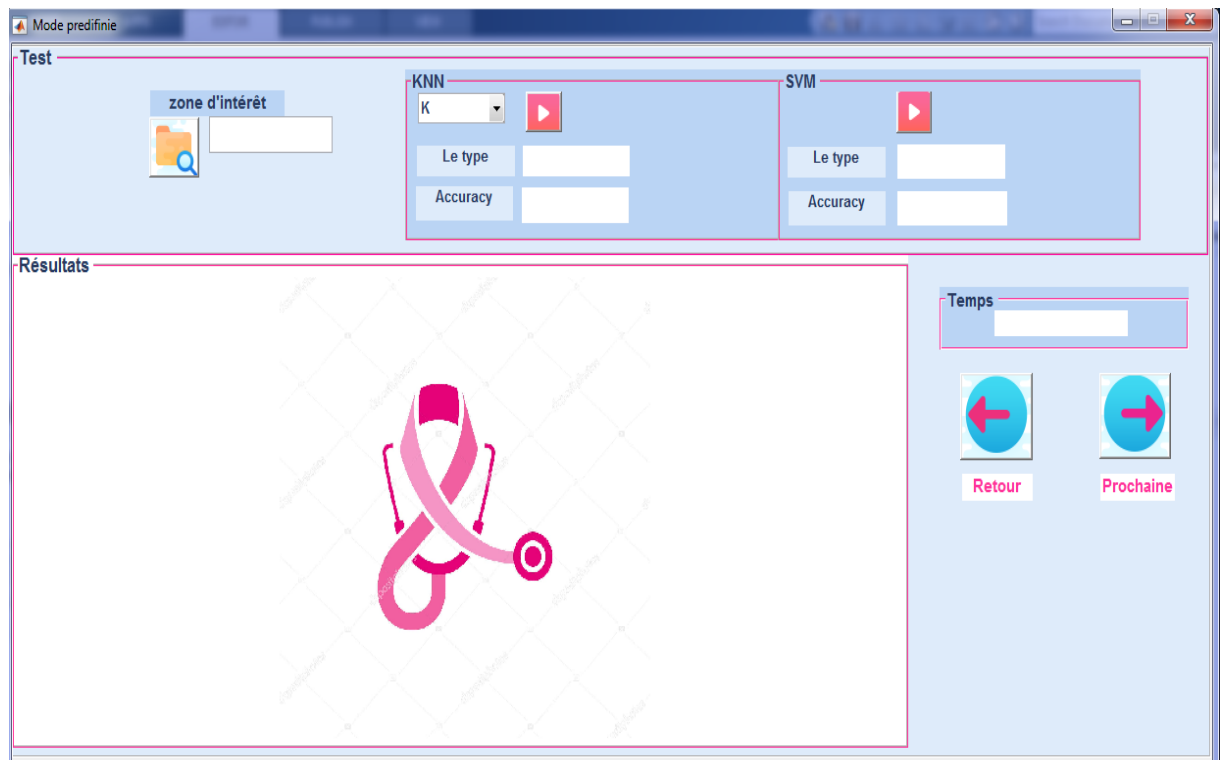


Fig V.12 Fenêtre du processus de décision médicale issue de l'utilisation de la sortie de Mammo-ZipfLBP.

Comme mentionné sur les fig V.13, 3 régions d'intérêt parmi les 5 renvoyées par le processus de recherche portent une pathologie maligne aidant le radiologue à diagnostiquer la région d'intérêt requête comme tumeur de type malin ce qui est vraiment le cas. En outre, le système Mammo-ZipfLBP offre un affichage des voisins appartenant à la même classe pathologique de la région d'intérêt requête.

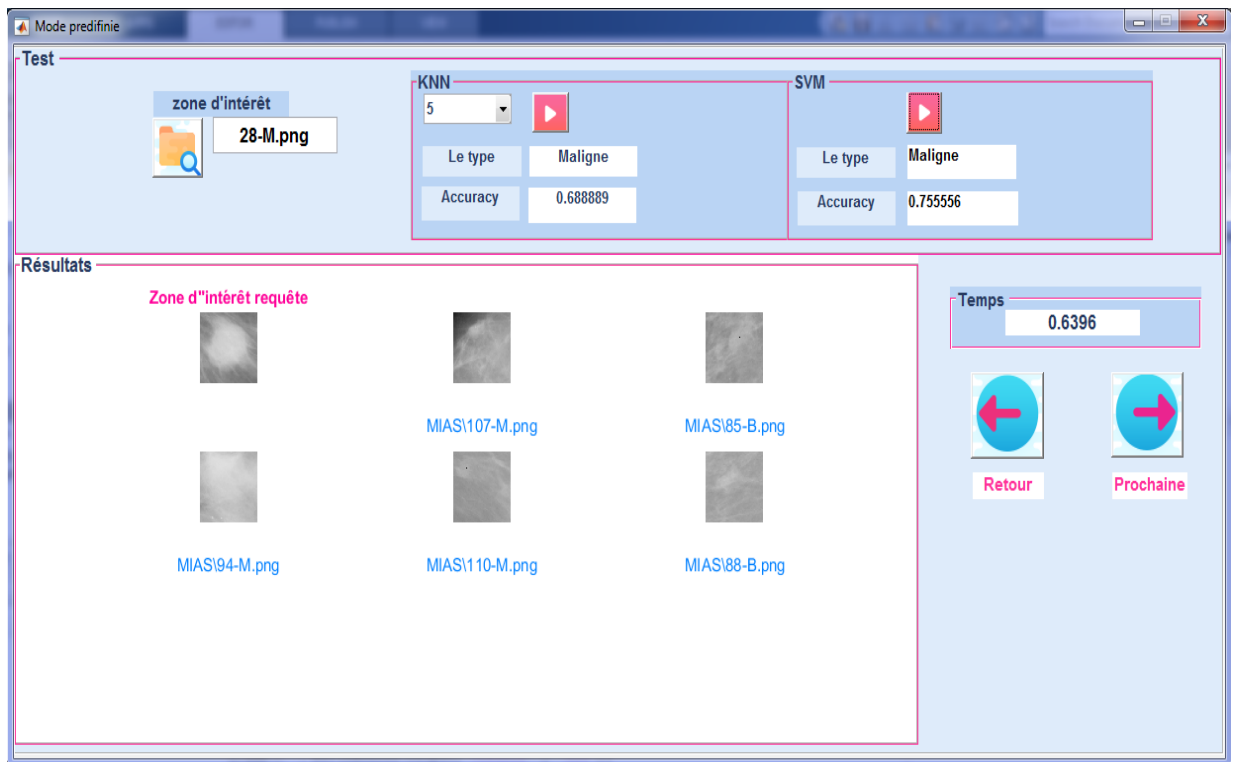


Fig V.13 Processus de décision médicale issue de l'utilisation de la sortie de Mammo-ZipfLBP avec K=5 d'une requête maligne.

V.9 Conclusion

Dans ce chapitre, nous avons développé un système d'indexation et de recherche de mammographies par le contenu «Mammo-ZipfLBP» basé sur une technique Template-matching à travers le classifieur des K plus proches voisins ainsi que celui des machines à vecteur de support. Nous avons obtenu une précision de classification assez bonne de l'ordre de 69%, ceci par les K-PPV et de l'ordre de 76% pour les SVM.

Conclusion générale

L'imagerie médicale s'est étendu de manière exponentielle en quantité du fait de l'accroissement du nombre de visites médicales par jour, ceci a rendu de l'utilisation efficace des images médicales un défi à élever pour les experts du domaine. En effet, la conception de systèmes performants qui permettent un stockage et une recherche de ces images était d'un grand besoin.

Dans plusieurs cas diagnostiqués par les radiologues, nous distinguons des ambiguïtés : l'anomalie focale pour un radiologue peut s'agir d'un tissu mammaire sein vis-à-vis un autre radiologue. En effet, la vision par ordinateur a surgit pour réduire les erreurs effectués dans la mammographie de dépistage en développant des systèmes d'aide au diagnostic médical assistés par ordinateur (CAD).

L'analyse de texture a prouvé efficacement la distinction entre sujets normaux et ceux pathologiques durant le processus de classification. En effet, le descripteur de la forme caractérise la géométrie des objets où deux tumeurs de natures différentes peuvent avoir le même contour tandis que la texture caractérise le tissu mammaire qui est une tache complexe pour le radiologue du fait que la variation du contraste dans certaines mammographies ne fut pas visible à l'œil nu.

Dans ce mémoire, nous avons traité l'aide au diagnostic médical du cancer du sein assisté par ordinateur, plus précisément, nous nous sommes appuyée sur l'indexation et la recherche d'images par contenu où une fusion entre deux techniques d'analyse de la texture des régions d'intérêt segmentées dans les mammographies, consistant en les lois puissance Zipf et Zipf inverse qui ont prouvé leur pertinence dans l'analyse d'image et les LBP (Local Binary Pattern) qui ont prouvé à leur tour la distinction entre les tissus pathologiques et ceux normaux. De plus, nous avons effectué une sélection des descripteurs les plus discriminants par les algorithmes génétiques.

Notre système «Mammo-ZipfLBP» se base sur l'approche Template matching qui affecte la région d'intérêt en cours d'analyse à la classe pathologique la majoritairement renvoyée par le processus de recherche où nous avons obtenu une détermination de classification assez bonne de l'ordre de 69% par l'algorithme des K plus proches voisins, et une précision de classification de l'ordre de 76% par l'algorithme des machines à vecteur de support.

Malgré les travaux riches des chercheurs dans le domaine de la détection et le diagnostic médical du cancer du sein assisté par ordinateur, néanmoins, ceci présente une deuxième opinion au radiologue qui demeure irremplaçable pour l'établissement du diagnostic final.

Perspective

Comme futurs travaux, nous exploiterons de nouvelles techniques d'analyse de la texture, de sélection des descripteurs les plus discriminants ainsi que de classification aussi nous essaierons de faire la classification des images non incluses dans la base de données.

Bibliographie

- [1] Mendoza, F., & Lu, R. (2015). Basics of image analysis. In *Hyperspectral Imaging Technology in Food and Agriculture* (pp. 9-56). Springer, New York, NY.
- [2] Prats-Montalbán, J. M., De Juan, A., & Ferrer, A. (2011). Multivariate image analysis: a review with applications. *Chemometrics and intelligent laboratory systems*, 107(1), 1-23.
- [3] Aguilera, J. M., & Stanley, D. W. (1999). *Microstructural principles of food processing and engineering*. Springer Science & Business Media.
- [4] Edmond Boyer (2013). *Informatique Visuelle (Visual Computing) – introduction à l'analyse d'images*.
- [5] P. Strumillo (2011) *IMAGE ANALYSIS and segmantation*.
- [6] BADEREDDINE, B. (2018). Analyse de la texture des images mammaires par une fusion des lois de Zipf et des SFTA pour la classification des tumeurs mammaires via l'analyse en composantes principales.
- [7] Dipert, B., & Shoham, A. (2012). Eye, robot: Embedded vision, the next big thing in digital signal processing. *IEEE Solid-State Circuits Magazine*, 4(2), 26-29.
- [8] R. Lepage (2015) *La vision par ordinateur*.
- [9] Di Cataldo, S., & Ficarra, E. (2017). Mining textural knowledge in biological images: Applications, methods and trends. *Computational and structural biotechnology journal*, 15, 56-67.
- [10] Machado, B. B., & Junior, J. F. R. (2017, July). Texture analysis using complex system models: fractal dimension, swarm systems and non-linear diffusion. In *30º Concurso de Teses Dissertações (CTD_2017)* (Vol. 30, No. 1/2017). SBC.
- [11] Ribas, L. C., Gonçalves, W. N., & Bruno, O. M. (2016). Texture analysis using deterministic partially self-avoiding walk with thresholds. *arXiv preprint arXiv:1611.08629*.
- [12] Linda Shapiro, George Stockman (Mars 2000) *Computer Vision* PP 235-236.
- [13] Hamoud, M. (2014). *Indexation et segmentation d'images basées loi de Zipf et Zipf inverse*. Thèse de doctorat. University Badji Mokhtar-Annaba.
- [14] Meriem, H., & Farida, M. H. (2011, November). Detection of a Region of Interest in the Images Based on Zipf Laws. In *2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems* (pp. 416-421). IEEE.
- [15] L. Ralaivola (2007) *Projet informatique : Algorithme des k-plus-proches-voisins*.

- [16] PAULUS DOMINIQUE, MAMBOURG FRANCOISE, BONNEUX LUC (2005) Dépistage du cancer du sein.
- [17] Institut National du Cancer- Boulogne-Billancourt Cedex (juin 2010) Tumeur maligne, affection maligne du tissu lymphatique ou hématopoïétique -Cancer du sein.
- [18] Cheikhrouhou, I. (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein (Doctoral dissertation).
- [19] Edson Damasceno Carvalho, Antonio Oseas de Carvalho Filho ,a, Alcilene Dalília de Sousa, Aristófanés Corrêa Silvab, Marcelo Gattassc (2018) Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity.
- [22] Institut national du cancer. DÉPISTAGE DES CANCERS DU SEIN, édition 2017 [en ligne]. Boulogne-Billancourt, Institut national du cancer, 2017 [consulté le 20/07/2017].
- [24] Puddu, M., & Tafforeau, J. (2005). Opportunité de dépistage du cancer du sein chez les femmes de 40 à 49 ans. Nr, 1, 1-267.
- [25] AKAOUI - BEDE – BOUSSOUKAYA (1 juin 2017) Loi de Zipf - université de lille1.
- [26] D'Orsi, C. J., Bassett, L. W., Berg, W. A., Feig, S. A., Jackson, V. P., Kopans, D. B., Linver, M. N., Mendelson, E. B., Moss, L. J., and Sickles., E. A. (2003). American college of radiology (breast imaging reporting and data system). Troisième édition française réalisée par SFR (Société Française de Radiologie).
- [27] Kramer, O. (2017). Genetic algorithm essentials (Vol. 679). Springer.
- [28] Hong, S. S., Lee, W., & Han, M. M. (2015). The feature selection method based on genetic algorithm for efficient of text clustering and text classification. Int. J. Advance Soft Compu. Appl, 7(1), 2074-8523.
- [29] Alexander Bozniak M (2007) , Entraînement de CAD's en mammographie , mémoire de licence , Université de Mons-Hainaut Faculté des Sciences Institut d'Informatique P28
- [30] Caron, Y., Makris, P., & Vincent, N. (2007). Use of power law models in detecting region of interest. Pattern recognition, 40(9), 2521-2529.
- [31] Nicole VINCENT, septembre (2004), Contribution de la loi de Zipf à l'analyse d'images, Université François Rabelais.
- [32] Faust, O., Acharya, U. R., Meiburger, K. M., Molinari, F., Koh, J. E., Yeong, C. H., ... & Ng, K. H. (2018). Comparative assessment of texture features for the identification of cancer in ultrasound images: a review. Biocybernetics and Biomedical Engineering, 38(2), 275-296.

- [33] Carvalho, E. D., de Carvalho Filho, A. O., de Sousa, A. D., Silva, A. C., & Gattass, M. (2018). Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity. *Computers & Electrical Engineering*, 67, 210-222.
- [34] Chang, R. F., & Lo, C. M. (2014, June). Computer-aided diagnosis for B-mode, elastography and automated breast ultrasound. In *International Workshop on Digital Mammography*(pp. 9-15). Springer, Cham.
- [35] Fujita, H., Hara, T., & Muramatsu, C. (Eds.). (2014). *Breast Imaging: 12th International Workshop, IWDM 2014, Gifu City, Japan, June 29-July 2, 2014, Proceedings* (Vol. 8539). Springer.
- [36] Rangayyan, R. M., Oloumi, F., & Nguyen, T. M. (2010). Fractal analysis of contours of breast masses in mammograms via the power spectra of their signatures. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*(pp. 6737-6740). IEEE.
- [37] Li, H., Giger, M. L., Olopade, O. I., & Lan, L. (2007). Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment. *Academic radiology*, 14(5), 513-521.
- [38] Beheshti, S. M. A., Noubari, H. A., Fatemizadeh, E., & Khalili, M. (2016). Classification of abnormalities in mammograms by new asymmetric fractal features. *Biocybernetics and Biomedical Engineering*, 36(1), 56-65.
- [39] Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical image analysis*, 37, 114-128.

Liens internet

- [20] https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_sein_pm.
- [21] http://www.ikonet.com/fr/ledictionnairevisuel/static/qc/types_cancers_sein_diagnostic
- [23] <https://www.cancer.be/les-cancers/types-de-cancers/cancer-du-sein/causes>.

Résumé: Le diagnostic assisté par ordinateur du cancer du sein devient de plus en plus une nécessité vu la croissance exponentielle du nombre de mammographies effectuées chaque année. L'analyse des images mammaires s'effectue au biais de méthodes linéaires et non-linéaires. En effet, les premiers sont simples à appliquer mais limitées alors le choix de la non-linéarité est d'un intérêt crucial dans le développement de puissants outils d'analyse d'image.

Dans ce sens, notre problématique de recherche consiste à appliquer les lois puissance : Zipf et Zipf inverse à l'analyse des images mammaires. En effet, les lois de Zipf caractérisent la complexité structurelle de la texture d'image par la modélisation de la répartition statistique de la fréquence d'apparition des motifs selon une distribution en lois puissance. Par ailleurs, nous avons réalisé une fusion des descripteurs obtenus avec ceux générés suite à l'analyse des images mammaires par les LBP (Local Binary Patterns) où une étape de sélection des descripteurs les plus discriminants a été réalisée par les algorithmes génétiques.

De ce fait, nous avons proposé un système d'indexation et de recherche des mammographies par le contenu (CBMIIR) qui augmente la performance du diagnostic assisté par ordinateur au niveau de l'étape de la présentation du diagnostic aux radiologues. Effectivement, ces derniers sont plus familiarisés avec un diagnostic basé sur des cas renvoyé par une classification basée template-matching, diagnostiqués préalablement, similaires au cas en cours d'analyse plutôt que le résultat abstraits généré par un classifieur. L'évaluation des travaux proposés a donné des performances élevées.

Mot clé: Analyse d'image, Vision par ordinateur, Loi de Zipf, Loi de Zipf inverse, Classification, Indexation, LBP, Recherche par le contenu, Template matching, Algorithmes Génétiques.

Summary: Computer aided diagnosis of breast cancer is becoming a necessity given the exponential growth in the number of the annual performed mammograms. Mammogram images analysis is carried out using linear and non-linear methods. Indeed, linear ones are simple to apply but limited, therefore, the choice of non-linearity is of crucial interest in the development of powerful image analysis tools. In this sense, our research problem is to apply the power laws: Zipf and inverse Zipf to mammogram images analysis. Indeed, Zipf's laws characterize the structural complexity of the image texture by modeling the statistical distribution of patterns frequency of appearance as power law distribution. Furthermore, we have performed a fusion of the obtained descriptors with those generated once analyzing mammogram images with LBP (Local Binary Patterns) where a step of most discriminating descriptors selection was performed by the genetic algorithms. Thus, we have proposed a Content-based mammogram image indexing and retrieval system (CBMIIR) that boosts the performance of a computer aided diagnosis (CADx) at the stage of providing the diagnostic to radiologists. Indeed, radiologists are more familiar with diagnosis decision based upon case-adaptive classification where similar known cases, to the one under analysis, are retrieved and displayed through a classification based template-matching; rather than the abstract result generated by a classifier. The evaluation of the proposed approach gave high performances.

Key words: Image analysis, Computer vision, Zipf's law, inverse Zipf's law, Local Binary Patterns, Classification, Indexing, Retrieval by content, Genetic algorithms, Template-matching.

ملخص إن التشخيص بواسطة الحاسوب لسرطان الثدي أصبح ضروريا نظرا للتزايد الهائل في عدد التصوير الشعاعي الثدي خلال السنة. إن تحليل العديد من الصور الثديية يتم بطريقتين : خطية وغير خطية الطريقة حيث الأولى سهلة التطبيق و لكن ذات محدودية إن فإن اختيار الطريقة الغير خطية بعد ذو أهمية جسمية في تطوير وسائل تحليل الصور الثديية .

في هذا الإطار تكمن المشكلة في البحث وتتطلب تطبيق قانوني " القوة " قانون زيبي اف وقانون زيبي اف عكسي من أجل تحليل الصور الثديية. فهات القوانين تصنف النسيج البنوي المعقد للصورة بواسطة النمذجة أي نمذجة التوزيع الإحصائي لتردد ظهور الأسباب وفق قوانين القوة . بالإضافة إلى ذلك قمنا بإجراء دمج للواصفات التي تم الحصول عليها مع تلك التي تم إنشاؤها بعد تحليل الصور الثديية من قبل (الأنماط الثنائية المحلية) LBP ثم تنفيذ خطوة اختيار أكثر الواصفات المميزة بواسطة الخوارزمية الجينية. في هذه الحالة قمنا باقتراح " نظام الفهرسة " والبحث عن الصور الثديية عبر المحتوى (CBMIIR) الذي يزيد من أداء التشخيص المحقق بمساعدة الحاسوب في مرحلة تقديم هذا التشخيص لأطباء الأشعة . طبعاً هؤلاء الأطباء المختصون في الإشعاع يتوافقون مع تشخيص مؤسس على نظام ترتيب الصور وهو أكثر تشابه للترتيب للصور الثديية وقد أعطى تقييم العمل المقترح أداء جيداً متجاوزاً نهج مختلفة مقترحة سابقاً.

الكلمات البحث : تحليل الصور الرؤية عبر الكمبيوتر قانون زيبي اف قانون زيبي اف معكوس تصنيف الفهرسة مطابقة القالب , البحث عن طريق المحتوى الأنماط الثنائية المحلية , الخوارزميات الجينية.