

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE

LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ECHAHID HAMMA LAKHDAR D'EL-OUED

FACULTÉ DES SCIENCES EXACTES

DEPARTEMENT D'INFORMATIQUE



MEMOIRE DE FIN D'ETUDE
MASTER ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Informatique

Spécialité: Systèmes Distribués et Intelligence Artificiel (SDIA)

THEME

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

Présenté par:

BADEREDDINE Belkis

DOUDI Hayet

Soutenu le xx Juin 2018

Devant le jury composé de :

Dr Hamoud Meriem

MC (B)

Encadreuse

M.

MAA l'Université d'El Oued

Président

M.

MAA l'Université d'El Oued

Examineur

Année universitaire: 2017/2018

Remerciements

Louange à Dieu qui nous a donné la force, le courage, et l'espoir nécessaire Pour accomplir ce travail et surmonter l'ensemble des difficultés.

Nous adressons notre plus vifs remerciements à Dr. Meriem HAMOUD pour sa confiance et son encadrement efficace nous reconnaissons la gentillesse et la spontanéité avec lesquelles elle a bien voulu diriger ce travail.

Nous tenons à remercier particulièrement les membres du jury qui nous ont fait

l'honneur d'accepter de juger notre travail .

Notre remerciement s'adresse également à nos professeurs de tout le cycle universitaire.

Nous remercions tous ceux qui nous ont aidé de près ou de loin pour élaborer ce travail.



Abstract

Medical images play a vital role in the diagnosis of patients, so the detection and classification of tumour is an important and delicate spot for the early diagnosis of the breast cancer.

In this work we applied a the laws powers Zipf and Zipf inverts besides the SFTA (Segmentation-based Fractal Texture Analysis), for the analysis of breast images to extract the descriptions textural from each technique and then integrate we then apply method the principal component analysis (PCA). We used the method Template-Matching based on the classification by the method of k-nearest neighbors (k-NN) to search for the most similar images in mammography to be chosen and classification of the pathological class of these images (malignant, benign and normal) using support vector machines (SVM) too. To assist the radiologists in making the final decision.

Keywords

Medical images, Classification, Laws powers Zipf and Zipf inverts, SFTA , PCA, Template-Matching, K-NN, Mammography, Pathological class, SVM.

نبذة مختصرة

تلعب الصور الطبية دورًا حيويًا في تشخيص المرضى ، لذا فإن اكتشاف الورم وتصنيفه هو موضع هام ودقيق للتشخيص المبكر لسرطان الثدي.

في هذا العمل قمنا بتطبيق قوانين القوى Zipf و Zipf مقلوب إلى جانب SFTA (تحليل التركيب الفركتري القائم على تقسيم) ، لتحليل صور الثدي لاستخراج الأوصاف النصية من كل تقنية ثم دمجها ثم تطبيق طريقة تحليل المكون الأساسي (PCA). استخدمنا طريقة مطابقة القوالب على أساس التصنيف باستخدام طريقة ك أقرب الجيران (k-NN) للبحث عن الصور الأكثر تشابهاً في التصوير الشعاعي للثدي الذي يتم اختياره وتصنيف الطبقة المرضية لهذه الصور (الخبثية و الحميدة ، طبيعى) باستخدام آلات متجه الدعم (SVM) أيضاً. لمساعدة أطباء الأشعة في اتخاذ القرار النهائي.

الكلمات المفتاحية

الصور الطبية-التصنيف- قوانين القوى Zipf و Zipf مقلوب - SFTA - PCA - مطابقة القوالب - K-NN - تصوير الثدي بالأشعة - الطبقة المرضية - SVM .

TABLE DES MATIÈRES

Remerciement

Abstract

ملخص

Table des matières

Liste des figures

Liste des tableaux

Liste des acronymes

Introduction générale.....1

Chapitre I .Analyse d'image et vision par ordinateur.....3

I.1. Introduction.....3

I.2. Lien entre l'analyse d'image et la vision par ordinateur.....4

I.3. Analyse d'image.....4

I.3.1. Analyse de bas niveau d'image.....5

I.3.2. Analyse d'intermédiaire niveau d'image.....5

I.3.3. Analyse de haut niveau d'image.....5

I.3.4. Analyse de prise de décision d'image.....6

I.4. L'imagerie médicale.....6

I.5. Analyse de la texture d'une image.....6

I.5.1 Définition de la texture.....7

I.5.2 Les méthodes d'analyse de la texture.....7

I.5.3 Les descripteurs de la texture.....11

I.6. La vision par ordinateur.....12

I.7. Conclusion.....14

Chapitre II. Les Lois puissance.....15

II.1. Introduction.....15

II.2 Définition d'une loi puissance.....16

II.3 Les principales lois puissance.....	17
II.3.1 La loi de Pareto.....	17
II.3.2 La loi de Zipf.....	18
II.3.3 La loi de Zipf inverse.....	20
II.4 Application des lois puissance en analyse d'images.....	20
II.4.1 Les lois de Zipf et de Zipf inverse en analyse d'image.....	20
II.4.2 Codage de l'image.....	21
II.4.3 Représentation graphique des lois de Zipf et de Zipf inverse.....	23
II.5. Conclusion.....	25
Chapitre III. Aide au diagnostic médical du cancer du sein assisté par ordinateur.....	26
III.1. Introduction.....	26
III.2. Le cancer du sein.....	27
III.3. Le dépistage du cancer du sein.....	28
III.3.1. La mammographie.....	28
III.4. Détection et classification assistées par ordinateur (CAdE/CADx) des tumeurs dans la mammographie.....	30
III.4.1. Détection des tumeurs assistée par ordinateur (CAdE).....	31
III.4.2. Diagnostic des tumeurs assisté par ordinateur (CADx).....	31
III.5. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.....	32
III.6. L'apport 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) se basant sur la sortie d'un classifieur.....	34
III.7. Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et les SFTA pour la classification des zones d'intérêt des mammographies.....	35
III.7.1. Analyse et caractérisation de la texture des zones d'intérêt par les lois de Zipf et de Zipf inverse.....	36
III.7.2. Analyse et caractérisation de la texture des zones d'intérêt par les SFTA.....	40
III.7.3. Bilan.....	41
III.8. Conclusion.....	42
Chapitre IV. Conception et Implémentation.....	43
IV.1. Introduction.....	43
IV.2. Problématique.....	44
IV.3. Classification.....	45
IV.3.1. Les méthodes de classification non-supervisées.....	46
IV.3.2. Les méthodes de classification supervisées.....	46
IV.4. Classification basée Template-matching.....	46
IV.5. Conception.....	47

IV.5.1. Mammo_ZS : le système proposé pour l’indexation et la recherche des mammographies par le contenu.....	47
IV.5.1.1. Les Etapes suivies et accomplies pour la réalisation du système «Mammo_ZS».....	47
IV.5.1.2. L’architecture globale du système «Mammo_ZS»	51
IV.6. Implémentation.....	53
IV.6.1. L’environnement de développement.....	53
IV.6.2. La base de données MIAS.....	53
IV.6.3.Evaluation de performance.....	53
IV.6.4. Présentation de quelques vues du système «Mammo_ZS».....	55
IV.7.Conclusion.....	58
Conclusion Générale.....	59
Bibliographie.....	61
Résumé	

LISTE DES FIGURES

Figure I.1 Calcul de la matrice de cooccurrence normalisée avec $d = 1$ et $\varnothing = 0$	8
Figure II.1 Représentation d'une loi puissance dans un repère linéaire.....	16
Figure II.2 Représentation d'une loi puissance dans un repère bi-logarithmique.....	17
Figure II.3 La fréquence de la distribution des revenus.....	18
Figure II.4 Tracé de rang/fréquence sur une échelle normale dans la figure 1a, tandis que la figure 1b montre le même graphique de rang / fréquence sur une échelle de log-log.....	19
Figure II.5 Codage d'un motif 3x3 d'une image par la méthode des 9 classes où en (a) nous présentons le motif original, et en (b) le motif codé par les 9 classes.	22
Figure II.6 Codage d'un motif 3x3 d'une image par la méthode des rangs généraux où en (a) nous présentons le motif original, et en (b) le motif codé par les rangs généraux	23
Figure II.7 Courbes de Zipf et de Zipf inverse obtenues suite au codage de l'image Lena par le codage des rangs généraux.....	24
Figure III.1 Vue latérale de l'anatomie et de la structure du sein.....	28
Figure III.2 Processus de mammographie numérique.....	30
Figure III.3 Deux types de vue de l'imagerie mammaire.....	30
Figure III.4 Architecture d'un système d'indexation et de recherche des images par le contenu (CBIR). 33	
Figure III.5 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux.....	37
Figure IV.1 Le processus d'obtention des descripteurs finaux.....	49
Figure IV.2 Architecture global du système proposée.....	51
Figure IV.3 Fenêtre Login du système «Mammo_ZS».	55
Figure IV.4 Interface principal de du système «Mammo_ZS».	56
Figure IV.5 Fenêtre des détails de réalisation de ce travail.....	56
Figure IV.6 Processus de décision médicale issue de l'utilisation de la sortie de Mammo_ZS par la considération de $k=5$	57

LISTE DES TABLEAUX

Table. IV.1 Variation de la précision de classification du système «Mammo_ZS» pour divers valeurs des k plus proche voisins.....	54
Table. IV.2 Performance du système «Mammo_ZS» avec et sans (ACP).....	55
Table IV.3 L'approche proposée vis-à-vis des approches.....	58

LISTE DES ACRONYMES

ACP	Analyse en Composantes Principales
bROI	Region Of Interest binaires
CC	Cranio-Caudal
CADe	Computer Aided Detection
CADx	Computer Aided Diagnosis
CBMIIR	Content Based Mammogram Image Indexing and Retrieval
CBIR	Content Based Image Retrieval
INSP	Institut National de Santé Publique
K-PPV	k- plus proches voisins
L'IRM	Imagerie par Résonance Magnétique
MLO	Médio-Latéral Oblique
MIAS	Mammographic Image Analysis Society
ROI	Region Of Interest
SFTA	Segmentation-based Fractal Texture Analysis
SVM	Support Vector Machine
TEP	Tomographie par émission de positons
TEMP	Tomographie par émission monophotonique
TFD	Transformée de Fourier Discrète
TCD	Transformée de Cosinus Discrète
TTBD	Two-Threshold Binary Decomposition
MATLAB	(matrix laboratory)

INTRODUCTION GÉNÉRALE

Le cancer du sein est l'une des raisons les plus largement reconnues de l'accroissement du taux de la mortalité chez les femmes. Pour palier à cette incidence, la détection précoce et le traitement sont les pivots de l'exploration de la pathologie mammaire. Ceci, à travers la mammographie de dépistage. A cet effet, le diagnostic assisté par ordinateur semble être une solution adéquate en développant des systèmes d'aide au diagnostic médical du cancer du sein destinés aux radiologues qui prennent le diagnostic fournie par la machine comme un deuxième avis pour l'élaboration de leur diagnostic final.

L'analyse détaillée de la mammographie est une tâche cruciale pour les radiologues, sans doute, le diagnostic assisté par ordinateur diminuera les biopsies suggérées par les radiologues lorsque la lecture visuelle ne suffit pas dans le cas par exemple où les radiologues rapportent plusieurs interprétations pour la même image mammaire. Par conséquent, à l'heure actuelle, la vision par ordinateur aide les experts médicaux au diagnostic, en devenant un outil incontournable dans le domaine médical.

La problématique qui a été posée pour l'élaboration de ce travail est l'analyse des images mammaires par des méthodes non linéaires qui répondent mieux à la structure complexe que présente une image au lieu d'appliquer de simples méthodes linéaires qui montrent vite leurs limites. Dans ce sens, nous nous intéressons aux lois puissance : Zipf et Zipf inverse pour l'analyse de la texture des images mammaires. Plus précisément, nous proposons une fusion des lois de Zipf et de Zipf inverse avec l'approche des SFTA dans le but 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.

L'objectif 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, l'unique différence 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 réside dans la présentation du diagnostic. Nous visons à renforcer la performance du diagnostic assisté par ordinateur (CADx) basé sur un résultat numérique généré par un classifieur et dont les radiologues ne sont pas familiarisés avec un diagnostic à base de cas traités et stockés préalablement dans une base de donnée à qui les radiologues exprime une familiarisation.

Ce travail de recherche est structuré comme suit :

Chapitre I : Analyse d'image et vision par ordinateur

Dans le chapitre I, nous présenterons les notions et concepts de base de l'analyse d'image et de la vision par ordinateur. Ensuite nous détaillerons l'analyse de la texture dans les images.

Chapitre II : Les lois puissance

Dans le chapitre II, nous présenterons un ensemble de lois puissances en se focalisant sur deux lois : Zipf et Zipf inverse en analyse d'image.

Chapitre III : Aide au diagnostic médical du cancer du sein assisté par ordinateur

Dans le chapitre III, nous présenterons les techniques assistés par ordinateur d'aide au diagnostic médical du cancer du sein. Ensuite nous détaillerons notre approche contributive dans ce mémoire.

Chapitre IV : Conception et Implémentation

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

Nous achèverons ce travail par une conclusion générale, en résumant notre contribution et en posant les perspectives de notre travail.

CHAPITRE I

I.1. Introduction

Nous évoquons dans le premier chapitre les notions et concepts de base de l'analyse d'image et de la vision par ordinateur. En outre, nous évoquons l'un des descripteurs les plus pertinents en vision par ordinateur consistant en la texture. En effet, nous nous focalisons durant notre travail sur la caractérisation de la texture par une fusion des lois puissance : Zipf et Zipf inverse et les SFTA (Segmentation-based Fractal Texture Analysis) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur.

I.2. Lien entre l'analyse d'image et la vision par ordinateur

L'analyse d'image est l'extraction d'informations significatives à partir d'une ou plusieurs images en se basant sur un nombre restreint de caractéristiques saillantes. Les informations extraites sont des attributs qui sont sélectionnés pour décrire la forme. L'analyse d'images permet la reconnaissance des éléments et peut également aller jusqu'à la localisation et l'identification des différents objets présents dans une image. Comprendre automatiquement le contenu d'une image est un des enjeux majeurs actuels, en effet, de nombreux chercheurs ont contribué à développer la nouvelle discipline qu'était la vision artificielle.

La vision par ordinateur vise à gagner une compréhension de haut niveau à partir d'images de natures très diverses provenant d'appareils photos, de caméras, de télescopes, de radars ou encore de scanners médicaux. L'ultime but est d'automatiser les tâches que le système visuel humain peut faire et implique le développement d'une base théorique et algorithmique pour atteindre la compréhension visuelle automatique [1]. En d'autres termes, l'objectif ultime de la vision par ordinateur est la compréhension de l'image, la capacité non seulement de récupérer la structure de l'image, mais aussi de savoir ce qu'elle représente. La vision par ordinateur est à la fois théorique (par exemple formulation de flux optique) et expérimentale (par exemple, réplique de modèle, réglage des paramètres) [2].

I.3. Analyse d'image

Les images peuvent être vues comme un ensemble de pixels caractérisés par leur intensité lumineuse variable, représentant une scène concrète ou abstraite. Le pixel est le terme le plus utilisé pour désigner les éléments d'une image numérique [3].

Le terme analyse d'image désigne l'analyse informatique des images numérisées et permet des mesures à partir de ces images. Actuellement, l'analyse d'image comprend de nombreuses classes d'outils, tels que le traitement du signal, l'intelligence artificielle et la stéréologie. Dans le domaine de la vision par ordinateur, les images sont généralement acquises par une caméra vidéo, puis numérisées par un ordinateur via une carte d'acquisition. Un système d'analyse d'image de vision par ordinateur peut être représenté comme un pipeline de traitement de données composé d'une série d'algorithmes. De tels algorithmes sont regroupés en étapes ou niveaux fonctionnels.

Les algorithmes de chaque niveau partagent certaines caractéristiques. Dipert et al. Identifient trois niveaux principaux dans [4], à savoir:

1. Acquisition d'image et optimisation.
2. Convertir des pixels en objets.
3. Analyse et reconnaissance des objets.

I.3.1. Analyse de bas niveau d'image

Le traitement de bas niveau ou de pixel correspond au niveau fonctionnel qui améliore la qualité ou transforme l'image d'entrée en une image qui convient à d'autres niveaux de traitement d'objet. Des exemples d'algorithmes exécutés à ce niveau sont le filtrage du bruit et l'amélioration du contraste pour améliorer la qualité, le recadrage ou la mise à l'échelle afin d'adapter les dimensions de l'image, la suppression de l'arrière-plan pour éliminer les informations indésirables. D'autres types de filtres, tels que les filtres à gradient pour produire des images de bords, peuvent également être nécessaires pour transformer l'image pour les niveaux suivants.

I.3.2. Analyse d'intermédiaire niveau d'image

Niveau intermédiaire ou traitement d'objet est le niveau fonctionnel qui est responsable du passage d'une représentation purement basée sur l'image à une représentation d'objet. Son but est d'identifier les régions potentielles d'intérêt pour un traitement ultérieur de haut niveau. Ainsi, il s'agit d'algorithmes de segmentation, ainsi que des algorithmes d'extraction de caractéristiques ou de descripteurs.

I.3.3. Analyse de haut niveau d'image

Le traitement de haut niveau ou la reconnaissance d'objet consiste à classer un objet en tant que membre d'une classe, citons l'exemple d'un visage humain en détection de visage ou en tant qu'instance particulière d'un objet, s'appuie sur des connaissances relatives au contenu de l'image. Des méthodes d'intelligence artificielle sont souvent appliquées dans le domaine de l'analyse et l'interprétation de l'image. Effectivement, la phase de vision par ordinateur de haut niveau s'appuie sur l'imitation de la cognition humaine en se basant sur l'information contenue dans l'image.

I.3.4. Analyse de prise de décision d'image

Le contrôle d'application ou le niveau de prise de décision constitue le dernier niveau de traitement. Comme son nom l'indique, ce niveau s'appuie sur les informations acquises lors des étapes précédentes pour prendre des décisions de haut niveau [5].

I.4. L'imagerie médicale

Les images médicales sont devenues une pierre clé dans le diagnostic et le suivi de presque toutes les maladies. Ces images offrent aux médecins la possibilité de participer à l'examen et au traitement des patients. La nature, le nombre et la résolution des images médicales ne cessent de croître grâce aux progrès constants des technologies d'acquisition d'images. Outre les radiographies, quatre grandes modalités d'imagerie sont couramment utilisées à l'hôpital : le scanner (tomodensitométrie par rayons X), l'IRM (imagerie par résonance magnétique), l'échographie (imagerie ultrasonore) et la scinti-graphie TEP ou TEMP (médecine nucléaire). Les images produites par ces quatre modalités sont volumiques: elles fournissent en chaque point du corps humain des informations mesurées dans un petit élément de volume appelé voxel (volume élément), l'extension volumique du pixel (image élément) [6]. Tout ce qui s'applique aux images numériques s'applique aux images médicales à la différence que les erreurs d'algorithmes dans le domaine médical peuvent conduire à la vie du patient.

I.5. Analyse de la texture d'une image

La texture forme l'une des caractéristiques fondamentales des objets, avec la couleur et la forme. Dans de nombreuses applications d'analyse d'image, il s'agit souvent de la seule qualité exploitable des objets [7] [8].

- L'analyse de texture consiste souvent à explorer des éléments de texture puis à révéler les règles de composition ou la nature répétitive de ces éléments de base.
- L'analyse de texture fait référence à une classe de procédures et de modèles mathématiques qui caractérisent les variations spatiales dans l'imagerie comme moyen d'extraire l'information.
- L'analyse de texture a trouvé la plus grande variété d'applications, qui comprennent la segmentation de différentes zones anatomiques, la différenciation entre sujets normaux et pathologiques ainsi que la classification d'un grand nombre de conditions pathologiques.

- L'analyse de texture tente de formaliser un concept intrinsèquement informel, c'est-à-dire l'apparence et la sensation des textures visuelles dans une image.

I.5.1 Définition de la texture

Dans le domaine du traitement d'image, il n'existe pas de définition conventionnelle de la texture. Les définitions mathématiques proposées peuvent être trop générales et donc imprécises et peuvent aussi être trop détaillées et donc ne couvrent pas tous les cas de texture envisageables [9]. Beaucoup de chercheurs ont essayé de donner une définition à "la texture" en prenant en considération certaine perspective de leur nature.

Dans [10], les chercheurs ont affirmé qu'en raison de la difficulté de comprendre et de modéliser le système de perception visuelle humaine, ils ne connaissent toujours pas, exactement, la nature de l'information que l'être humain capte pour analyser et interpréter une image qu'il faudrait fournir à un système de vision artificielle pour qu'il puisse analyser et comprendre une image, comme l'être humain l'effectue. Haralick considère une texture comme un "phénomène de zone organisée" qui peut être décomposé en "primitives" ayant des distributions spatiales spécifiques [11]. Haidekker a défini la texture comme "une variation locale systématique des valeurs d'image" [12]. Petrou a déclaré que "la caractéristique la plus importante de la texture est qu'elle dépend de l'échelle et que "différents types de texture sont visibles à différentes échelles" [13].

Dans [14], les auteurs affirment qu'une texture est une zone de l'image qui présente certaines caractéristiques d'homogénéité qui la font apparaître comme une zone unique ou comme étant un ensemble de primitives de taille et de forme variables, présentant une organisation spatiale particulière.

I.5.2 Les méthodes d'analyse de la texture

Les méthodes d'analyse de la texture peuvent être classées en six groupes mentionnés ci-dessous [15] :

- **Les méthodes structurelles**

Cette catégorie d'approches applique la définition de base de la texture comme une répétition régulière de sous-modèles ou de primitives. En effet, ils identifient d'abord de telles primitives, appelées aussi éléments de texture, puis calculent des descripteurs statistiques ou

morphologiques en supposant certaines règles de placement des primitives.

L'hypothèse d'un placement homogène des primitives est une limitation majeure. Bien que cette hypothèse soit généralement très bonne pour les textures artificielles, elle fut réfutée dans les images biologiques [10].

• Les méthodes statistiques

Elles représentent les textures par la distribution spatiale des pixels de niveau gris dans l'image [11]. La texture est une distribution spatiale non-déterministe des valeurs d'intensité, peut être représentée mathématiquement par un ensemble de statistiques du premier ou du second ordre:

- Les statistiques du premier ordre concernent la probabilité que des pixels individuels aient des valeurs d'intensité spécifiques. ces statistiques sont collectées à partir de l'histogramme d'intensité normalisé de l'image.
- Les statistiques du second ordre se rapportent à la vraisemblance conjointe de deux pixels aléatoires dans l'image ayant des paires spécifiques de valeurs d'intensité où la probabilité conjointe des paires de pixels est prise en compte. Un histogramme d'intensité de second ordre doit être calculé, appelé matrice de cooccurrence [16], qui est une matrice carrée où chaque élément en position (i, j) contient la probabilité pour une paire de pixels situés à une distance d et une direction h dans l'image pour avoir les niveaux d'intensité i et j, respectivement comme mentionné sur l'exemple de la figure I.1

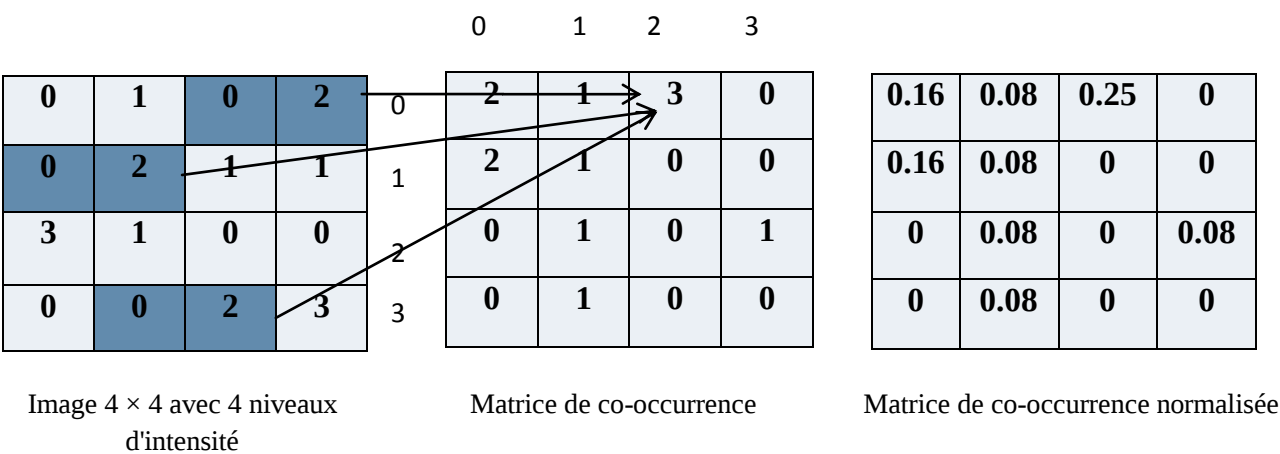


Figure I.1 Calcul de la matrice de cooccurrence normalisée avec d = 1 et ø = 0. Source [8]

D'un autre côté, ce type de descripteur a deux limitations majeures. Tout d'abord, la difficulté de régler l'orientation du dipôle (d , h) afin d'obtenir une discrimination optimale de la texture, qui peut être très dépendante de l'image. Deuxièmement, le manque d'invariance des descripteurs obtenus à la taille et la rotation. Par conséquent, les versions tournées ou mises à l'échelle de la même texture seront étiquetées différemment, ce qui entraînera des erreurs de classification [8].

- **Méthodes spectrales**

Inspirées par le fait que le système visuel humain décrit des images dans le domaine fréquentiel [17]. Elles convertissent l'image dans une nouvelle image basée sur les propriétés des fréquences spatiales de l'intensité des pixels [18]. Les descripteurs de texture sont généralement déduits à partir d'images filtrées, sur un certain nombre de domaines différents. Nous citons les plus utilisés:

- Filtres de domaine spatial. Les méthodes de domaine spatial naïf reposent sur des opérateurs de détection de contours simples (par exemple, Sobel), puis extraient la densité des arêtes dans l'image filtrée, en l'utilisant comme un descripteur de texture. Cette approche permet de distinguer les schémas grossiers des schémas fins, mais elle a de lourdes limitations pour gérer les textures irrégulières [8].

- Filtres de domaine de fréquence. L'analyse de fréquence peut être appliquée, soit au moyen de Transformée de Fourier Discrète (TFD) bidimensionnelle, soit par transformée de Cosinus Discrète (TCD), pour extraire les composantes de fréquence spatiale des images.

Nous affirmons que dans le domaine fréquentiel spatial, les propriétés de texture globales telles que la grossièreté, la granulation ou les motifs répétitifs peuvent être facilement identifiés. Les coefficients des transformées fournissent une représentation compacte de l'image originale où les motifs les plus discriminants sont soulignés. Dans la littérature, ils sont largement utilisés comme descripteurs de texture, soit tels quels, soit sous la forme de caractéristiques statistiques ou d'histogrammes de coefficients [8].

- Gabor et les transformées en ondelettes. A la différence de la TCD et la TFD, les ondelettes effectuent des décompositions spatiotemporelles où la base sinusoïdale est modulée avec des fonctions de fenêtres de formes différentes. La présence d'une fenêtre de largeur limitée permet généralement une meilleure localisation dans le domaine spatial par rapport aux

décompositions de Fourier traditionnelles, assurant les meilleures capacités de discrimination. De plus, différentes formes de fenêtres peuvent s'adapter à différents types de textures. [8]

Malgré que les filtres de Gabor présentent une résolution d'image conjointe dans les domaines spatial et fréquentiel, ceci dit ils ne décrivent pas des textures bien grossières; en effet, l'énergie de ces textures est concentrée dans les sous-images de fréquences les plus basses.

L'analyse par ondelettes est fondamentalement une analyse linéaire et souffre de résolutions uniformément mauvaises sur différentes échelles ainsi que de sa nature non adaptable aux données, du fait que la même base d'ondelettes est utilisée pour analyser toutes les données [17].

- **Méthodes basées sur un modèle**

Dans les méthodes basées sur un modèle, les paramètres estimés des modèles a priori assument le rôle de descripteurs de texture et peuvent être utilisés pour la synthèse de texture, la classification ou la segmentation. Les modèles les plus utilisés dans la littérature sont [8]:

- Modèles autorégressifs. Ils supposent une interaction locale directe entre les pixels de l'image, de sorte que l'intensité du pixel est une somme pondérée des intensités de pixel dans un voisinage du pixel et un bruit distribué de façon identique. Les paramètres du modèle sont représentés par le vecteur des poids.

- Champs aléatoires. La texture peut être considérée comme un échantillon fini d'un processus aléatoire bidimensionnel qui peut être décrit par ses paramètres statistiques.

Les champs aléatoires de Markov sont une généralisation multidimensionnelle des chaînes de Markov, définies en termes de probabilités conditionnelles associées aux quartiers spatiaux.

Par conséquent, la représentation et l'analyse des textures sont traduites en un problème d'inférence statistique. Le principal inconvénient de ces techniques est la charge de calcul due aux schémas itératifs d'optimisation de l'énergie.

- Fractales. Une fractale est un concept mathématique où un ensemble multi-échelle présente le même motif répété à chaque échelle, ce qui est un paradigme qui peut être facilement transféré à l'analyse de texture. En effet, les paramètres fractaux peuvent être considérés comme une mesure de l'irrégularité ou de l'hétérogénéité des arrangements spatiaux.

Donc, ces descripteurs peuvent être utilisés comme descripteurs de texture dans des problèmes de classification où les textures sont caractérisées par une forte irrégularité, comme dans les images histologiques des tissus cancéreux [8]. Par exemple, l'algorithme d'analyse de texture fractale basée sur la segmentation (SFTA). L'algorithme d'extraction consiste à décomposer l'image d'entrée en un ensemble d'images binaires à partir desquelles les dimensions fractales des bordures des régions sont calculées pour décrire les motifs de texture segmentés. Afin de décomposer l'image d'entrée.

- **Méthodes basées sur un agent**

Des méthodes basées sur des agents ont été proposées pour l'analyse de texture [18]. Ces méthodes utilisent un individu guidé par une règle déterministe ou stochastique, qui effectue des marches dans l'image et utilise les informations de la promenade comme vecteur de descripteurs. L'ultime avantage de cette catégorie par rapport aux autres méthodes littéraires est sa capacité à décrire efficacement les micro-textures.

- **Méthodes basées sur des graphes**

Ces méthodes représentent une image en tant que graphique, où chaque pixel est un sommet et les bords sont générés en ce qui concerne l'emplacement et l'intensité des pixels voisins, ou les réseaux neuronaux complexes sont appliqués pour la classification [15].

I.5.3 Les descripteurs de la texture

Un observateur humain décrit la texture de manière qualitative : fine, grossière, foncée, grainée, lisse... etc. Donc, une quantification de ces propriétés est obligatoire pour la création de descripteurs numériques mathématiques ayant le but d'identifier la texture précisément pour rendre la reconnaissance par machine possible.

Au cours des dernières décennies, les descripteurs sont devenus des outils principaux dans la communauté de vision par ordinateur et furent largement appliqués dans un grand nombre d'applications, citons, la représentation de l'image, la classification et la récupération des images, la reconnaissance et l'appariement des objets.

Une bonne classification des textures dépend de la présence de descripteurs discriminants et représentatifs dans l'image. Ainsi, la détection et l'extraction des descripteurs de l'image présentent des étapes essentielles pour ces applications.

Tout d'abord, nous détectons des régions d'intérêt, ensuite, pour chaque région détectée, une représentation vectorielle invariante (un vecteur descripteur) pour les données d'image autour des régions d'intérêt détectées est construite. Ces descripteurs extraits de l'image peuvent être issus du second ordre statistique, des modèles paramétriques, des coefficients obtenus à partir d'une transformation d'image, ou même une fusion de ces mesures.

Nous distinguons deux types de descripteurs; à savoir les descripteurs globales et les descripteurs locales. Nous mentionnons que les descripteurs issus de l'analyse de la texture ont une caractérisation globale de la texture, en effet, ils visent à décrire une image dans son ensemble et peuvent être interprétées comme une propriété particulière de l'image impliquant tous les pixels. La texture peut être trouvée dans toute l'image ou dans certaines régions de celle-ci.

Notons que la littérature s'accorde sur un point, l'existence de deux types de textures : les microtextures, composées de motifs dont l'échelle est de l'ordre de quelques pixels, et les macrotextures, généralement composées de motifs plus larges [19] [20].

I.6. La vision par ordinateur

Nous distinguons deux techniques d'analyse d'images, une à base de cortex visuel humain et l'autre par à base de l'ordinateur. En effet, la vision par ordinateur vise à reproduire l'effet de la vision humaine en percevant électroniquement une image.

Si nous abordons l'analyse d'image basée sur le cortex visuel humain, nous constatons qu'en tant qu'être humain, nous sommes d'excellents analystes d'images pour en extraire des caractéristiques de plus haut niveau. Ceci se manifeste dans de nombreuses applications telles que la médecine, la sécurité et les applications de télédétection et bien d'autres. Il est connu que le cortex visuel humain est un excellent appareil d'analyse d'images.

De l'autre part, les systèmes de vision par ordinateur envahissent les applications industrielles et scientifiques, notamment les systèmes de contrôle de robots et de véhicules autonomes, la vidéosurveillance et l'inspection, l'organisation de bases de données d'images et la modélisation objet.

L'ultime but est de remplacer l'observateur humain durant l'interprétation du contenu de l'image pour donner en sortie du système de la vision par ordinateur une représentation abstraite de l'aspect particulier de l'image ou de tout le contenu de l'image.

L'analyse de textures est une problématique importante et ancienne de l'analyse d'image et de la vision par ordinateur. La plupart des méthodes d'analyse de texture ont principalement été développées pour le traitement d'images 2D. En vision par ordinateur, l'enjeu majeur réside dans l'extraction de caractéristiques visuelles de la texture consistant en des descripteurs numériques pertinents et discriminants permettant la détection d'objets ou la recherche d'images par le contenu. En effet, l'analyse de la texture présente un grand intérêt en vision humaine et en psychophysique, cependant, les caractéristiques de texture peuvent ne pas être facilement perçues par la vision humaine. A cet effet, nous allons nous baser sur la caractérisation de la texture par une fusion des lois puissance : Zipf et Zipf inverse et les SFTA (Segmentation-based Fractal Texture Analysis) afin de bénéficier de leur apport complémentaire [21] [22] [23] dans un processus d'aide au diagnostic médical du cancer du sein assisté par ordinateur.

I.7. Conclusion

Le développement et l'avancée rapide de la technologie Internet et de l'informatisation nécessitent des techniques d'analyse et de traitement d'images très précises et puissantes, pour l'obtention d'informations parfaites, utiles et sémantiques à partir d'images afin de les utiliser dans de nombreuses applications scientifiques ainsi qu'industrielles.

Les ordinateurs permettent d'éviter l'interprétation subjective et fastidieuse de l'image. Plus important encore, la manipulation et l'interprétation informatisées des données fournissent une précision et une rapidité sans l'intégration d'erreurs humaines.

L'analyse et l'interprétation de l'image constituent l'outil principal de la vision par ordinateur en traitant des problèmes dont les humains résolvent à leur insu, nous nous focaliserons sur cet enjeu durant l'élaboration de ce mémoire. En effet, nous nous intéressons aux modèles de loi puissance de type: Zipf et Zipf inverse ainsi qu'aux SFTA pour la caractérisation de la texture des zones d'intérêt extraites à partir d'images mammaires pour donner un aide au diagnostic médical du cancer du sein assisté par ordinateur.

CHAPITRE II

II.1. Introduction

Les lois de puissance sont utilisées dans de nombreuses disciplines scientifiques. Les premières lois proviennent probablement des enquêtes de Pareto (1897) sur la répartition des revenus personnels. Il convient également de souligner la contribution significative du philologue George Kingsley Zipf (1949), consistant en la loi de Zipf et la loi de Zipf inverse.

Dans ce chapitre, une définition des lois puissance est présentée, spécialement les lois de Zipf et de Zipf inverse, puis nous évoquerons l'adaptation nécessaire pour les appliquer à notre domaine de recherche, à savoir l'analyse d'image et la vision par ordinateur.

II.2 Définition d'une loi puissance

Une loi de puissance est une relation dans laquelle un changement relatif d'une quantité donne lieu à un changement relatif proportionnel dans l'autre quantité, indépendamment de la taille initiale de ces quantités [24]. On peut définir une loi puissance comme étant une fonction associant deux quantités. Elle peut être de la forme suivante :

$$Y = ax^{-b} \quad (II.1)$$

Où **a** et **b** sont des constantes.

Sa représentation graphique dans un repère linéaire (x, y), est comme indiqué dans le graphique ci-dessous (figure II.1). La "courbe de puissance" est comme une asymptote sur les deux axes du graphique pour mettre en évidence la non-linéarité de la relation entre les deux variables.

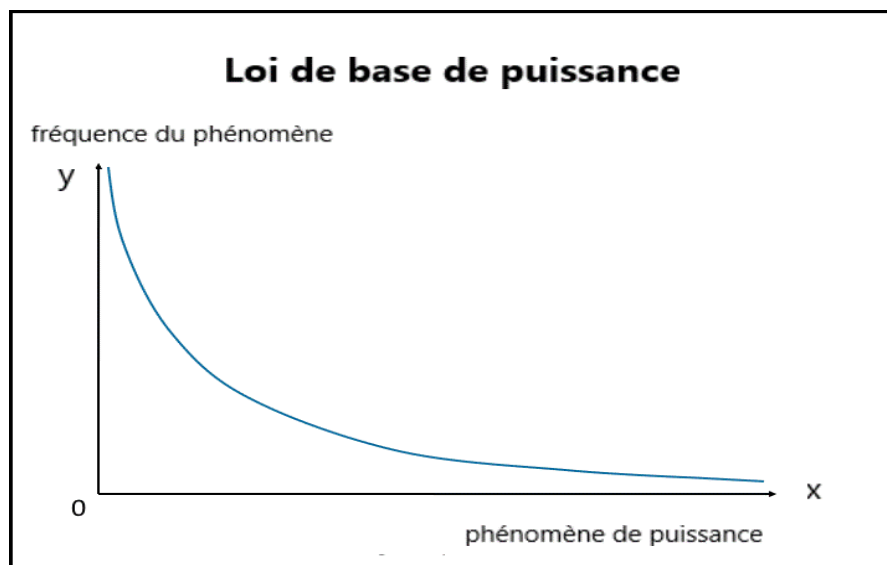


Figure II.1 Représentation d'une loi puissance dans un repère linéaire.

En général, les lois puissance sont représentées dans un repère bi-logarithme. En choisissant un tel repère, nous obtenons la représentation graphique d'une distribution en loi puissance sous la forme d'une droite. Nous avons donc $\ln y = \ln a - b \ln x$. Comme nous pouvons le constater sur la figure II.2 [25].

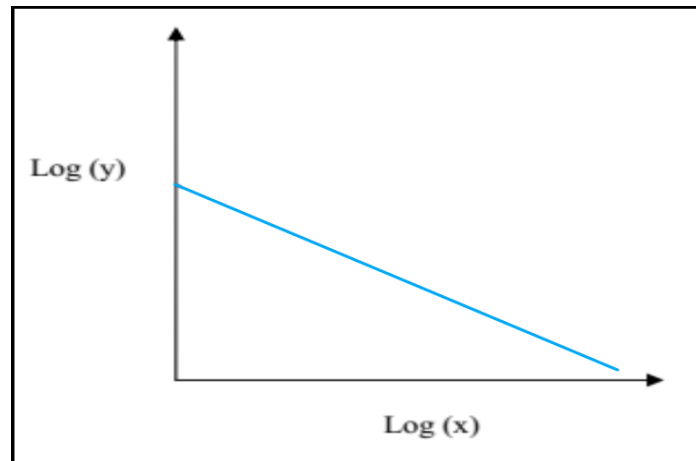


Figure II.2 Représentation d'une loi puissance dans un repère bi-logarithmique.

Les modèles de loi puissance les plus connus sont la loi de Zipf et la loi de Pareto. En effet, ces lois ont été appliquées dans plusieurs disciplines scientifiques telles que la physique, l'ingénierie, l'informatique, les géosciences, la biologie, l'économie, la linguistique et les sciences sociales. Cette omniprésence a favorisée de plus en plus l'intérêt de la recherche au cours de la dernière décennie.

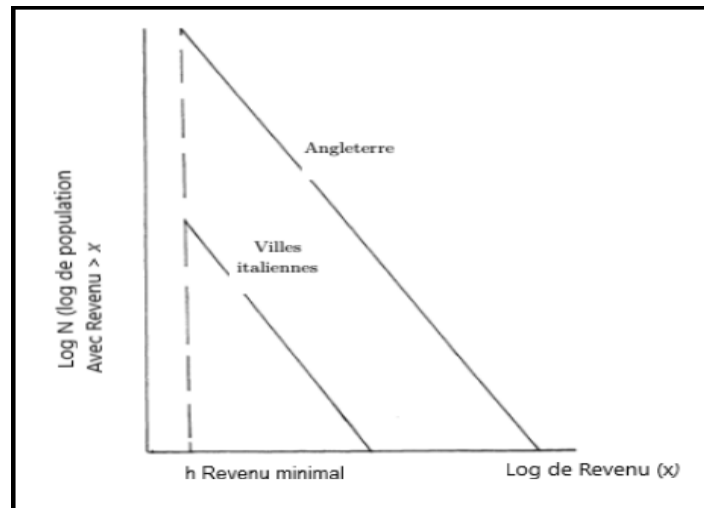
II.3 Les principales lois puissance

II.3.1 La loi de Pareto

Généralement, la présence de distributions de loi de puissance est un indice de la complexité sous-jacente d'un système, ceci présente un défi pour les scientifiques de modéliser et d'étudier le problème. C'est pourquoi la loi de Pareto ou encore la « loi des 80/20 » est l'une des premières loi puissance (1897) mise en place par l'économiste Wilfried Pareto [26]. En étudiant la répartition des revenus personnel des individus et de la richesse dans les principaux pays industrialisés, il a conclu que 80% des ressources appartenaient à seulement 20% de la population. Où il a constaté que cette répartition suivait une loi puissance. Le principe de Pareto indique que, pour de nombreux événements, environ 80% des effets proviennent de 20% des causes. Il base ses démonstrations à partir des variations de l'inégalité dans la distribution. En considérant des données pour la Grande Bretagne, plusieurs villes italiennes et certains Etats allemands, Pareto propose d'étudier la distribution cumulative des revenus des habitants de ces régions. Cette distribution est caractérisée par la relation [27]:

$$P(X > x) = \left(\frac{x}{x_{min}}\right)^{-k} \text{ avec } x \geq x_{min} \quad (II.2)$$

Où x représente le revenu des ménages, x_{min} le revenu minimal considéré et k un indice de hiérarchisation. La figure II.3 reproduit le schéma original de Pareto qui trace la courbe de la fréquence de distribution des revenus observés selon le niveau de revenu individuel [27].



Source : Pareto, 1895 [1965], schéma reproduit par Persky (1992). [28] [29]

Figure II.3 La fréquence de la distribution des revenus.

L'espérance mathématique du revenu est donnée par la formule suivante :

$$\bar{x} = x_{min} k / (k - 1) \quad (II.3)$$

La valeur de k mesure la pente de la distribution et mesure l'inégalité des revenus [25]. Plus généralement, la distribution 80/20 ou loi de Pareto peut être observée dans de nombreux autres domaines de la vie. Citons les exemples suivant : 20% des problèmes informatiques occupent 80% du temps du département informatique, 20% des travailleurs génèrent 80% des résultats, 20% des clients génèrent 80% des revenus.

II.3.2 La loi de Zipf

Auerbach a découvert en 1913 [30] que la taille de la ville est régie par une loi de puissance. Le linguiste américain [George Kingsley Zipf](#) a décrit une distribution de la loi de puissance en fréquence des mots en 1949 [31] (bien qu'Estroup l'ait remarqué en 1916). Zipf a enquêté davantage sur cette distribution, révélant que la même distribution en loi de puissance tient

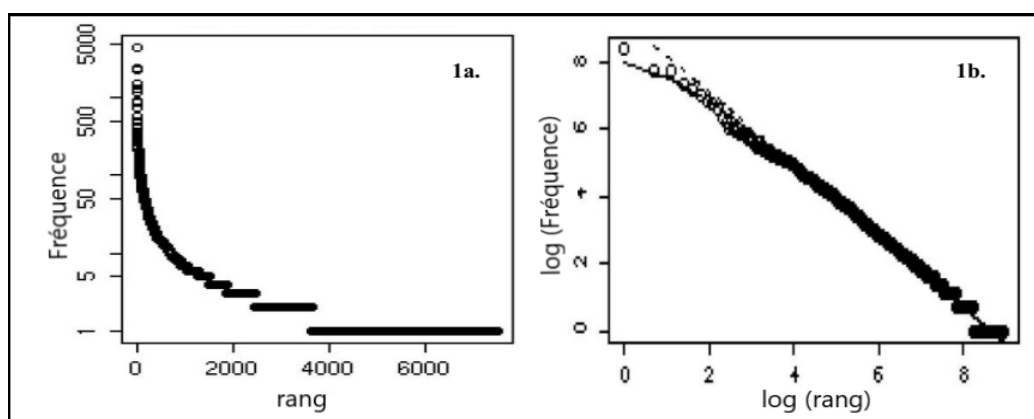
pour un grand nombre d'événements dans différents domaines, allant des tailles de tremblements de terre, revenu annuel des personnes, éruptions solaires, au nombre de citations reçues sur des documents. Sous la loi de Zipf (Zipf 1949) [31], les distributions empiriques des mots sont comme suit :

Relativement peu de mots sont utilisés fréquemment tandis que la plupart des mots se produisent rarement, beaucoup ne se produisant qu'une seule fois dans des échantillons de textes, même de grande taille. Plus précisément, la fréquence d'un mot tend à être approximativement inversement proportionnelle à son rang en fréquence. Soit f la fréquence du mot avec le rang de r dans un ensemble de N mots, alors [32]:

$$f = \frac{C}{r} \quad \text{où } C \text{ est une constante} \quad (\text{II.4})$$

En d'autres termes, le mot le plus fréquent (rang 1) dans un texte apparaîtra environ deux fois plus souvent dans un texte que le deuxième mot le plus fréquent (rang 2), trois fois plus souvent que le troisième mot le plus fréquent (rang 3), etc, le corpus lexical étudié étant arrangé par ordre décroissant d'occurrence de mot.

Cette distribution entraîne une dépendance linéaire entre rang et fréquence si les données sont tracées sur une échelle doublement logarithmique. Un exemple d'une telle distribution sur une échelle à la fois normale et doublement logarithmique est représenté sur la figure II.4 [33].



Source: Zipf's law in H.G. Well, *The War of the Worlds* (1989) (Baroni 2009). [34]

Figure II.4 Tracé de rang/fréquence sur une échelle normale dans la figure 1a, tandis que la figure 1b montre le même graphique de rang / fréquence sur une échelle de log-log.

II.3.3 La loi de Zipf inverse

Un autre comportement observé pour les textes écrits en langage naturel a été modélisé par une loi, appelée loi de Zipf inverse. Elle s'intéresse à la distribution de la fréquence des mots d'un texte. Zipf a découvert une loi de puissance qui est vérifiée uniquement pour les mots de faible fréquence. Contrairement à la loi de Zipf qui s'intéresse aux mots les plus fréquents, la loi de Zipf inverse permet donc d'étudier la répartition des mots les plus rares dans le texte : le nombre de mots distincts $I(f)$ de fréquence d'apparition f et la fréquence f sont liés par la relation [35] :

$$I(f) = Af^{-\gamma} \quad (\text{II.5})$$

Avec A et γ des constantes réelles. Donc, contrairement à la loi de Zipf précédemment définie qui décrit la répartition statistique des mots, la loi de Zipf inverse décrit la répartition statistique des fréquences des mots [25].

II.4 Application des lois puissance en analyse d'images

Les lois puissance, et en particulier les lois de Zipf et de Zipf inverse ont été étudiées et vérifiées dans de nombreux domaines. A cet effet, il serait intéressant de mentionner l'existence de répartitions en loi puissance dans un cas bidimensionnel, tel que l'image. L'analyse d'une image par les lois de Zipf désigne la modélisation de la fréquence d'apparition des motifs au sein de l'image selon des distributions en loi de puissance; donc, ces modèles de lois puissance caractérisent la complexité structurelle des textures dans les images. Suite à cette analyse, nous obtenons les courbes de Zipf et de Zipf inverse, cependant, ces courbes ne fournissent aucune mesure de texture qui peut être utilisée comme descripteurs, effectivement; nous extrayons l'information de la texture à travers un ensemble de descripteurs de texture modélisant l'information sur la distribution spatiale des niveaux de gris dans l'image.

II.4.1 Les lois de Zipf et de Zipf inverse en analyse d'image

- **La loi de Zipf**

Cette loi énonce que dans un ensemble de symboles topologiquement organisés, la distribution des numéros d'occurrence des n-uplets n'est pas organisée de manière aléatoire. Les n-uplets désignent les motifs dans le cas des images. En effet, Le motif $M_{2k+1}(i,j)$ de taille $2k+1$ centré sur le pixel $v(i,j)$ est défini comme suit pour $k > 1$:

$$M_{2k+1} = \begin{cases} v(i-k, j-k) \dots \dots v(i-k, j+k) \\ v(i, j) \\ v(i+k, j-k) \dots \dots v(i+k, j+k) \end{cases}$$

Les fréquences d'apparition des motifs $M_1, M_2 \dots M_n$ de l'image, notons $N_1, N_2 \dots N_n$, sont en relation avec le rang de ces motifs, lorsqu'ils sont triés par rapport à leur fréquence d'occurrence décroissante. La relation suivante peut être observée [36]:

$$N_{\sigma(i)} = k \cdot i^{-a} \quad (II.6)$$

Où $N_{\sigma(i)}$ représente la fréquence d'apparitions d'un motif de rang i , k et a sont des constantes.

- **La loi de Zipf inverse**

Cette loi décrit les fréquences d'apparition des n -uplets d'un ensemble de symboles 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 [37] [38]:

$$I(f) = A f^{-\gamma} \quad (II.7)$$

Où A et γ sont des constantes positives.

II.4.2 Codage de l'image

Une image est un ensemble de pixels organisés dans le plan sous la forme d'une matrice. Les symboles considérés pour la vérification de la loi de Zipf et de Zipf inverse, sont donc les niveaux de gris utilisés pour coder les pixels. Ces niveaux de gris sont codés sous la forme d'un entier codé sur 8 bits entre 0 et 255. Les n -uplets de symboles seront ici des motifs constitués de pixels adjacents de l'image, chaque motif est défini comme un masque 3×3 de pixels adjacents. Si les niveaux de gris étaient directement utilisés, la fréquence d'apparition de chaque motif particulier serait très faible en raison du grand nombre de motifs possibles, et la distribution des fréquences de motif ne serait pas statistiquement significative.

Nous devons donc définir un certain codage de l'image qui permet de réduire le nombre de motifs différents à une valeur acceptable pour que l'on puisse en dégager une information pertinente [39] [40].

- **Codage des 9 Classes**

Une technique simple réduisant le nombre de motifs différents susceptibles d'être présents dans une image, serait de diviser l'échelle des niveaux de gris [0, 255] en un nombre diminué de classes et d'affecter à chaque pixel la valeur de cette classe en fonction de sa valeur de luminance. Un exemple de codage de modèle avec cette méthode est donné sur la figure II.5. La classe $c(x, y)$ d'un pixel avec échelle de gris $g(x, y)$, où N est le nombre de des classes, est donné par la formule suivante [41] [25] :

$$c(x, y) = \text{int} \left[\frac{N * g(x, y)}{255} \right] \quad (\text{II.8})$$

La figure suivante montre un exemple de motif codé avec les 9 classes.

255	210	210	8	7	7
25	2	34	0	0	1
40	2	40	1	0	1
(a)			(b)		

Figure II.5 Codage d'un motif 3x3 d'une image par la méthode des 9 classes où en (a) nous présentons le motif original, et en (b) le motif codé par les 9 classes.

- **Codage des rangs généraux**

Le principe du codage des rangs généraux, est de respecter la vision d'une scène qui repose plus sur les variations locales de niveaux de gris, afin de mettre en évidence les différences relatives plutôt que les valeurs absolues. Ainsi, les valeurs des niveaux de gris sont remplacées par leurs rangs lorsqu'elles sont triées en fonction des valeurs de niveau de gris. La méthode affecte le même rang lorsque les niveaux de gris ont la même valeur. Nous affectons la valeur 0 au niveau de gris le plus bas, et on incrémente la valeur d'une unité quel que soit l'écart relatif entre deux niveaux de gris consécutifs.

Ce qui conduit à une très forte diminution du nombre de motifs différents possibles [39] [40]. La figure II.6 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)		

Figure II.6 Codage d'un motif 3x3 d'une image par la méthode des rangs généraux où en (a) nous présentons le motif original, et en (b) le motif codé par les rangs généraux.

II.4.3 Représentation graphique des lois de Zipf et de Zipf inverse

- **Construction de la courbe de Zipf**

L'algorithme de construction de la courbe de Zipf de l'image que nous visons à mettre en évidence ses propriétés texturales consiste à balayer l'image par un masque 3x3 en codant les motifs rencontrés puis à associer à chaque motif sa fréquence d'apparition dans l'image. Par la suite nous comptons le nombre d'occurrences de chacun des motifs.

Le tableau des motifs est initialement vide, ensuite nous rangeons dans ce tableau tous les motifs rencontrés dans l'image, en associant à chaque motif sa fréquence d'apparition. Pour chaque nouveau masque étudié, il est comparé successivement à chacun des motifs déjà présents dans le tableau. Si le motif est présent, nous incrémentons la fréquence d'apparition de l'unité. Mais, si le motif n'est pas dans le tableau nous le rajoutons au tableau en incrémentant sa fréquence d'apparition à 1.

A la fin et après avoir balayée l'image entière, nous trions la fréquence d'apparition des motifs par ordre décroissant. Ensuite nous traçons la courbe qui donne la fréquence d'apparition en fonction du rang dans un repère bi logarithmique [37] [38].

- **Construction de la courbe de Zipf inverse**

L'algorithme de construction de la courbe de Zipf inverse de l'image consiste à balayer l'image par un masque 3x3 en codant les motifs rencontrés puis à associer à chaque motif sa fréquence

d'apparition dans l'image. Ensuite, pour chaque valeur trouvée de la fréquence d'apparition, nous dénombrons les motifs différents ayant cette fréquence d'apparition.

Initialement, la fréquence cherchée est à 1, puis on parcourt séquentiellement le tableau des motifs afin de compter les motifs ayant la même fréquence que la fréquence courante. Nous réitérons ensuite l'algorithme en incrémentant à chaque fois la fréquence d'une unité jusqu'à ce que l'on ait atteint la fréquence maximale. A cet effet, nous traçons la représentation graphique dans un repère bi-logarithmique [37] [38].

La figure II.7 présente un exemple de courbe de Zipf et Zipf inverse relatives à l'image Lena en utilisant la méthode des rangs généraux pour le codage des motifs.

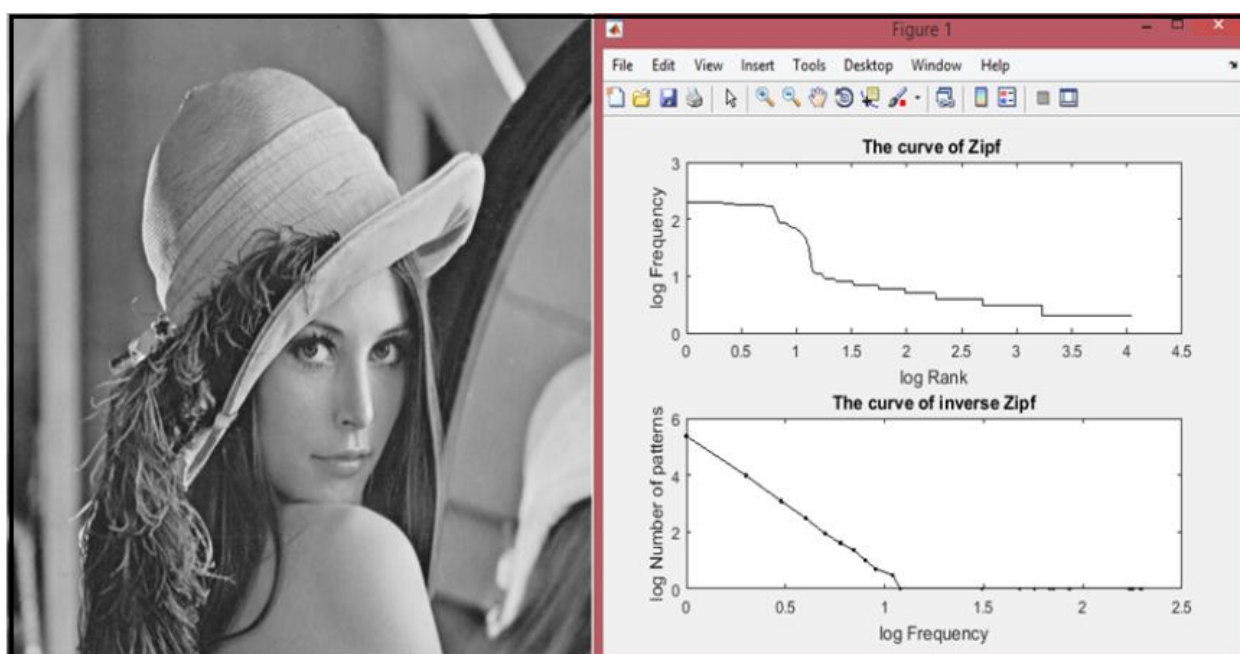


Figure II.7 Courbes de Zipf et de Zipf inverse obtenues suite au codage de l'image Lena par le codage des rangs généraux.

Rappelons que la loi de Zipf établie une relation entre les fréquences d'apparition des motifs de l'image et le rang de ces motifs dans une suite ordonnée, tandis que la loi de Zipf inverse établie une relation entre les fréquences d'apparitions des motifs et le nombre de motifs ayant la même fréquence d'apparition.

II.5. Conclusion

Les modèles loi puissance ont été utilisées non seulement en linguistique, mais aussi en économie, pour l'étude de la démographie urbaine, ou pour la modélisation du trafic Internet. En matière d'analyse d'images, les lois de Zipf et Zipf inverse sont adaptées au traitement des images et permettent d'évaluer la complexité structurelle d'une image par la répartition statistique des fréquences d'apparition des motifs.

Le codage est ensuite sélectionné pour ces motifs qui permettent de réduire le nombre de motifs différents dans l'image. Dans notre cas, nous allons appliquer le codage des rangs généraux vu qu'il est adéquat à notre problématique de la caractérisation de la texture au sein des images mammaires.

Ainsi, la technique d'analyse d'images basée sur les lois puissances : Zipf et Zipf inverse peut être considérée comme une alternative pertinente pour d'autres techniques d'analyse d'image purement statistiques ou purement linéaires évoquées dans le premier chapitre de ce mémoire.

Les lois de Zipf et de Zipf inverses se sont avérées efficaces pour analyser la structure complexe de l'image numérique, en particulier dans le domaine de l'imagerie médicale, où elles peuvent constituer un outil efficace d'aide au diagnostic médical du cancer du sein assisté par ordinateur par exemple. C'est ce que nous expliquerons dans le prochain chapitre où nous évoquerons le diagnostic médical du cancer du sein assisté par ordinateur à travers l'analyse numérique de la mammographie, pour aider les médecins à établir le diagnostic final.

CHAPITRE III

III.1. Introduction

Le cancer du sein est l'une des principales causes mondiales de décès chez les femmes, en effet, il provoque la mort de plus de 500 000 femmes chaque année dans le monde et représente à lui seul 30% de tous les nouveaux diagnostics de cancer chez les femmes. La détection précoce est fondamentale pour le traitement efficace du cancer du sein et la mammographie de dépistage est l'outil le plus couramment utilisé par la communauté médicale, qui augmente les chances de survie jusqu'à 90% [42].

L'analyse de l'image mammaire effectuée manuellement par un radiologue est très compliquée et implique principalement l'utilisation d'une mammographie numérique à faible contraste et provoque des erreurs de diagnostic par le radiologue dans 10% à 30% des cas, suite aux certaines structures masquées dans les images mammaires en raison de la fatigue visuelle du radiologue. Par conséquent, un système de diagnostic médical assisté par ordinateur est crucial pour aboutir à un diagnostic plus précis en identifiant et classifiant des masses bénignes et malignes pour les cas dans lesquels la détection et classification est très difficile à l'œil humain. Dans ce chapitre nous proposons une approche de caractérisation de la texture des images mammaires basée fusion des lois de puissance : Zipf, Zipf inverse et les SFTA pour la classification des zones d'intérêt extraites à partir des mammographies.

III.2. Le cancer du sein

L'organisation mondiale de la santé a classé le cancer du sein comme le type de cancer le plus courant chez les femmes. Selon les estimations, 1.671.149 nouveaux cas de cancer du sein ont été identifiés et 521.907 décès dus au cancer du sein sont survenus dans le monde en 2012 (GLOBOCAN 2012). Selon GLOBOCAN, le cancer du sein, représentant 25,1% de tous les cancers. Notons que l'incidence du cancer du sein dans les pays développés est plus élevée, tandis que la mortalité relative est plus élevée dans les pays moins développés. En Algérie, le cancer du sein est considéré comme étant le premier type de cancer, avec 11.000 cas recensés annuellement, sur les 45.000 nouveaux cas annuels de cancer, tous types confondus, selon les données de l'Institut national de Santé publique (INSP) de l'année de 2015. Divers facteurs ont été identifiés comme pouvant modifier le risque de cancer du sein chez les femmes, ces facteurs de risque se classent en deux catégories : les facteurs dits non génétiques, qui regroupent le style de vie, l'exposition environnementale, l'alimentation ; et les facteurs génétiques, observés dans le génome de l'individu.

Le cancer du sein est la conséquence de la croissance incontrôlée des cellules mammaires formant des masses. En effet, le sein féminin est principalement constitué de lobules (glandes productrices de lait), de canaux (passages de lait qui relient les lobules au mamelon), tissu gras et conjonctif entourant les canaux et les lobules, les vaisseaux sanguins et les vaisseaux lymphatiques comme mentionné sur la figure III.1. La plupart des cancers du sein ont leur origine dans les cellules des canaux, cellules des lobules. Le stade précoce du cancer canalaire est appelé in situ, ce qui implique que le cancer reste confiné aux canaux (carcinome canalaire in situ). Lorsque cela envahit le tissu adipeux environnant et se propage à d'autres organes, il est appelé invasive [43].

Les taux de survie varient considérablement dans le monde entier avec les taux les plus bas dans les pays en développement. Ces faibles taux peuvent s'expliquer principalement par l'absence de programmes de dépistage précoce ainsi que par un diagnostic inadéquat et une pénurie d'installations de traitement, ce qui entraîne une forte proportion de femmes présentant une maladie à un stade avancé. A cet effet, l'augmentation du taux de survie est en liaison directe avec la détection précoce grâce à un dépistage périodique.

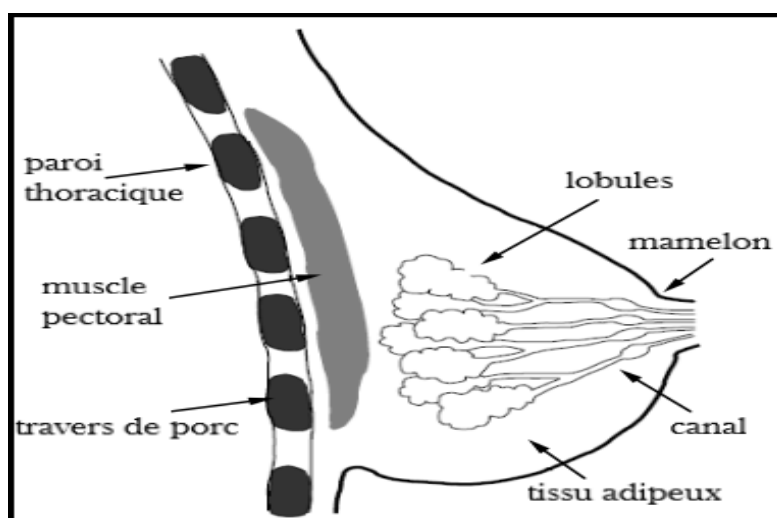


Figure III.1 Vue latérale de l'anatomie et de la structure du sein.

III.3. Le dépistage du cancer du sein

III.3.1. La mammographie

La mammographie est une technique radiographique qui utilise les rayons X à basse énergie pour visualiser le tissu mammaire et mettre en évidence les anomalies. Dans la mammographie, un faisceau de rayons X traverse chaque sein, où il est absorbé par le tissu en fonction de sa densité. Les rayons restants vont à un film photographique à travers le détecteur et produisent une image de niveau de gris après le développement. L'image de résultat est connue sous le nom de mammographie à base de film [43]. Deux projections d'image de chaque sein sont obtenues, vue Cranio-Caudal (CC) et vue Médio-Latéral Oblique (MLO). Il y a 3 types de mammographie, chacun étant réalisé pour des raisons différentes :

- **La mammographie de dépistage** : est un test de dépistage utilisé pour rechercher le cancer chez les femmes qui n'ont aucun symptôme de cancer du sein ou de problèmes mammaires. Les deux seins sont examinés lors d'une mammographie de dépistage. Il peut aider à trouver des masses ou des régions anormales de tissu mammaire qui peuvent être trop petites pour être ressenties à la main ou par palpation. La mammographie de dépistage surveille les changements au sein au fil du temps et peut aider à trouver le cancer du sein à un stade précoce [44].
- **La mammographie diagnostique** : est faite pour diagnostiquer un problème mammaire, comme une masse ou une région suspecte. Il est possible que vous observiez ce problème par vous-même et que vous le mentionniez à votre médecin, ou

qu'on détecte le trouble du sein lors d'un examen clinique des seins ou d'une mammographie de dépistage. La mammographie diagnostique prend plus de temps qu'une mammographie de dépistage. Des images plus détaillées et des vues de la poitrine sont prises sous des angles différents, pour regarder de plus près une zone dans la poitrine. La mammographie diagnostique est habituellement effectuée sur les deux seins afin que les médecins puissent comparer le tissu mammaire des deux seins [44].

- **Mammographie numérique (FFDM) :** La mammographie numérique permet d'obtenir une image électronique du sein plutôt qu'un film radiographique. Elle se déroule de la même façon qu'une mammographie standard sur film radiographique, mais les images sont stockées dans un ordinateur. Le processus de mammographie numérique est décrit schématiquement sur la figure III.2. Dans la pratique courante, il y a deux projections capturées pour chaque sein en mammographie: l'une est Cernio-Caudal (CC) et l'autre est Medio-Lateral Oblique (MLO) montré dans la figure III.3.

Parmi les nombreux avantages de la mammographie numérique, nous mentionnons:

- La possibilité de modifier les images pour obtenir une meilleure vue des zones anormales.
- Les images peuvent être transmises et stockées, donc nous pourrions les utiliser à des fins de comparaison future.
- Minimiser le temps mis pour développer le film avec une amélioration de la qualité de l'image.
- La tomo-synthèse et autres technologies de visualisation 3D.
- La réduction de la dose de rayonnement reçue par la patiente.
- La détection des tumeurs de taille médiocre.

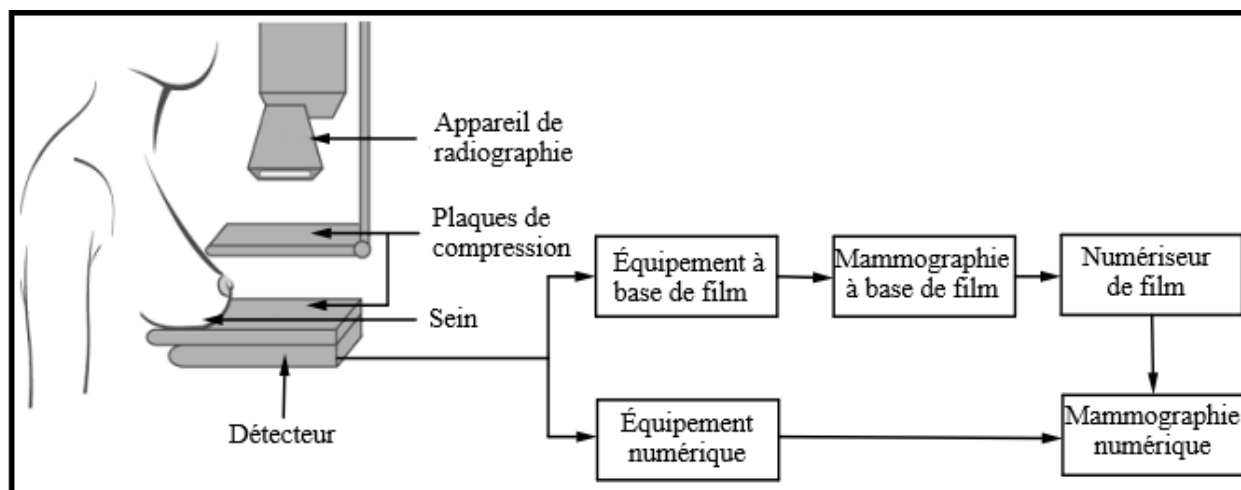


Figure III.2 Processus de mammographie numérique. Source [43]

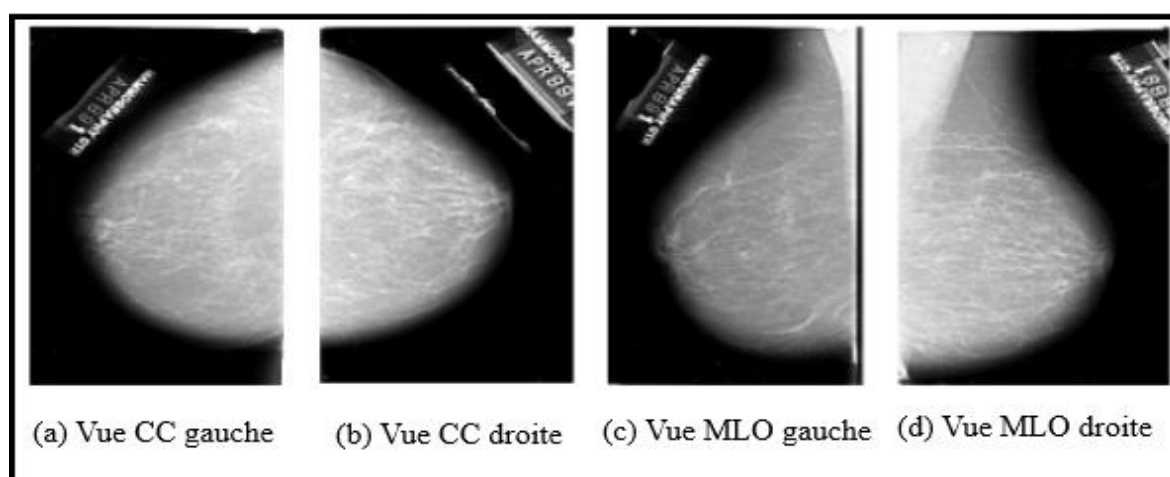


Figure III.3 Deux types de vue de l'imagerie mammaire. Source [43]

Notons que l'interprétation de la mammographie est une tâche répétitive qui nécessite une attention maximale pour éviter l'interprétation erronée. Par conséquent, il y a eu lieu l'apparition des systèmes de détection ou de diagnostic assisté par ordinateur (CAD) désignant une méthode extrêmement populaire et efficace qui analyse les mammographies numériques à l'aide de techniques de traitement d'image et de reconnaissance de formes.

III.4. Détection et classification assistées par ordinateur (CAdE/CADx) des tumeurs dans la mammographie

Un système de détection et de diagnostic assisté par ordinateur est une classe de systèmes informatiques qui visent à aider à la détection et / ou au diagnostic de maladies grâce à une «seconde opinion» ou « un deuxième avis » [45].

Le but des systèmes de détection et de diagnostic assisté par ordinateur est d'améliorer la précision des radiologues en réduisant le temps d'interprétation des images médicales. Ces systèmes sont classés 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). Les systèmes de détection des tumeurs assistée par ordinateur (CADE) sont des systèmes orientés à localiser les lésions dans les images médicales. De l'autre part, les systèmes de diagnostic des tumeurs assisté par ordinateur (CADx) effectuent la caractérisation des lésions, par exemple, la distinction entre les tumeurs bénignes et malignes.

III.4.1. Détection des tumeurs assistée par ordinateur (CADE)

Le but des systèmes de détection des tumeurs assistée par ordinateur (CADE) est de détecter les anomalies dans les tissus des mammifères, tels que les micro-assemblages agrégés et les masses.

La première étape de ces systèmes est la détection de régions suspectes. Les algorithmes les plus courants pour identifier les zones d'intérêt (ROI : region of interest) sont basés sur l'étude des pixels, le principal avantage de ces méthodes est leur implémentation simple, alors que leur inconvénient majeur réside dans leur processus de calcul intensif. Dans les techniques de détection basées sur les zones, les régions d'intérêt sont extraites par des techniques de segmentation, qui tiennent compte de la morphologie et de la taille des masses, elles ont une complexité de calcul inférieure à celle des méthodes basées sur les pixels. Tandis que la deuxième étape des algorithmes de détection de masse est la classification des zones d'intérêt (ROI) en tissu normal ou tissu avec lésion [46].

III.4.2. Diagnostic des tumeurs assisté par ordinateur (CADx)

Le but des systèmes de diagnostic ou de classification des tumeurs assistés par ordinateur (CADx) est de distinguer les lésions suspectes comme étant bénignes ou malignes afin de réduire le nombre de recommandations biopsiques sur les lésions bénignes. Une zone ou une tumeur est initialement identifiée par un radiologue ou un ordinateur (CADE), le système de diagnostic assisté par ordinateur (CADx) implique l'intégration de diverse opération de traitements d'images, où la segmentation d'image, l'extraction de caractéristiques, la sélection des caractéristiques les plus discriminants ainsi que la classification sont essentielles.

La segmentation est la base d'un système de diagnostic assisté par ordinateur (CADx) du fait que la probabilité de malignité dépend de la marge des lésions ou de la forme qui caractérise la géométrie des objets ou bien de la texture qui caractérise le tissu mammaire.

Les tâches de diagnostic sont conçues en extrayant les caractéristiques correspondantes des masses et des calcifications [46].

La section clé des systèmes de détection et de diagnostic des tumeurs assistés par ordinateur est l'étape de classification considérée comme le cœur de la méthode. En effet, la classification est une technique d'apprentissage automatique pour analyser les caractéristiques spatiales et les organiser en catégories souhaitables. Citons l'exemple du classifieur des K- plus proches voisins (K-PPV) qui se base sur une approche classique pour classifier les objets en fonction d'un échantillon d'apprentissage dans l'espace des caractéristiques [46].

Nous devons préciser que, malgré les travaux fructueux des chercheurs dans le domaine de la détection et le diagnostic médical du cancer du sein assisté par ordinateur (Comme [46] et [47]), cependant le radiologue demeure irremplaçable et le diagnostic final n'est spécifique qu'à ce dernier.

III.5. Indexation et recherche des mammographies par le contenu (CBMIIR) pour l'aide au diagnostic médical du cancer du sein assisté par ordinateur

Aujourd'hui, les services de soins de santé s'appuient fortement sur diverses données d'imagerie médicale qui ont augmenté de manière remarquable en quantité en raison de l'augmentation rapide du nombre de visites médicales par jour. Dans ce sens, les chercheurs se sont beaucoup intéressés au développement de techniques structurées pour gérer d'énormes bases de données d'images médicales afin de rendre l'accès, la recherche et la récupération efficaces.

Pour palier au problème du nombre très élevé des images mammaires, l'approche d'indexation et de recherche d'image par le contenu est diffusée visant le 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 d'accéder efficacement aux mammographies archivées souhaitées, pathologiquement similaires, à la mammographie en cours d'évaluation [37].

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. Dans 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 par le calcul d'une distance. Enfin, l'image cible sera obtenue à partir des images récupérées [43] [48]. L'architecture du système d'indexation et de recherche des images par le contenu est présentée sur la figure III.4.

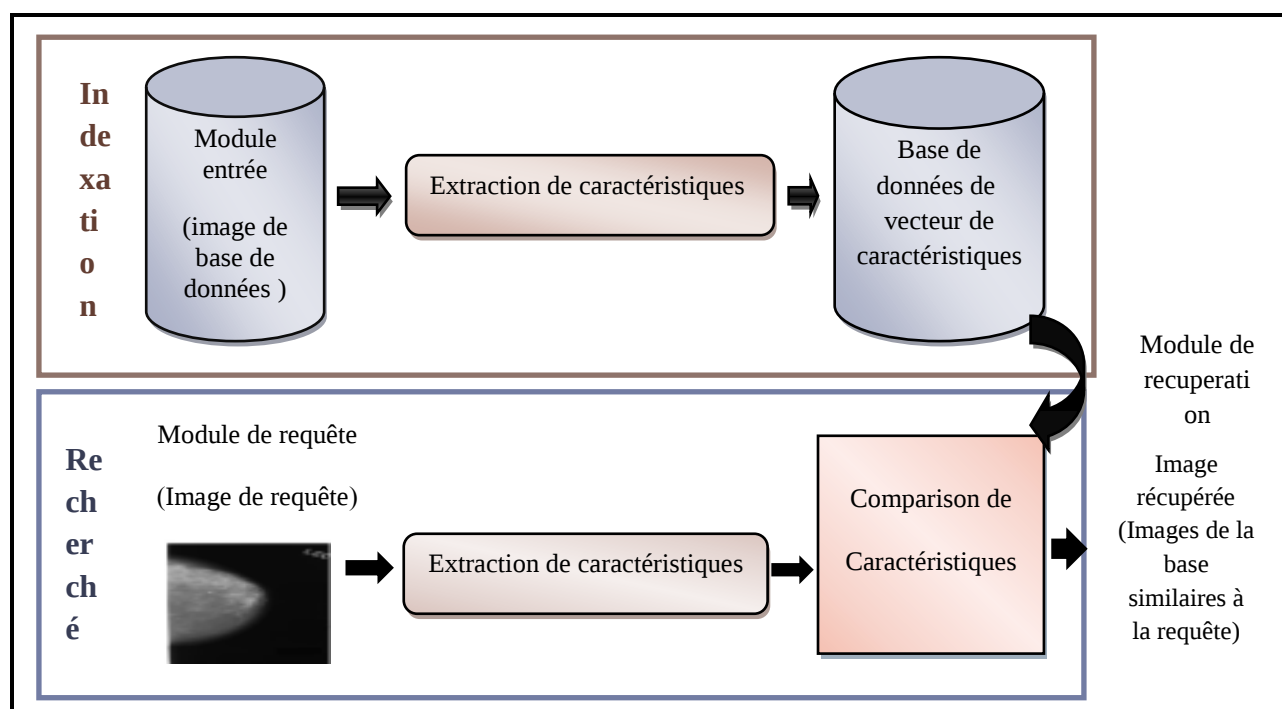


Figure III.4 Architecture d'un système d'indexation et de recherche des images par le contenu (CBIR).

La spécificité de ces méthodes est qu'elles se basent uniquement sur la comparaison du contenu visuel des images, sans nécessairement intégrer des aspects sémantiques. La base de données d'entités donc doit être constituée de caractéristiques de bas niveaux telles que des informations sur la couleur, les formes ou la texture.

III.6. L'apport 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) se basant sur la sortie d'un classifieur

Les systèmes de diagnostic assisté par ordinateur (CADx) offrent une aide substantielle au processus de prise de décision des radiologistes. Le but de tels systèmes est de réduire l'effort requis pour le diagnostic d'une lésion. La méthodologie principale derrière les systèmes utilisés implique l'extraction de caractéristiques des lésions et l'application d'un classificateur pour obtenir une discrimination entre les masses bénignes et malignes ainsi que le tissu sein. Cependant, les systèmes de CADx souffrent d'un taux plus élevé de faux positifs et de faux négatifs, qui produisent une quantité accrue de biopsies inutiles. A cet effet, il est nécessaire de développer des systèmes de CADx avec une précision améliorée pour aider les radiologistes dans une détection précise et une meilleure planification du traitement. Ces erreurs sont principalement causées par les performances relativement faibles des systèmes de CADe en détection de masse.

Une autre tendance dans cette direction concerne les systèmes d'indexation et recherche des images mammaires par le contenu (CBMIIR : Content Based Mammogram Image Indexing and Retrieval) offrant une décision dans le contexte d'images récupérées similaires à celles stockées dans une base de données de référence, fournissant ainsi aux radiologues une aide visuelle justifiant les résultats produits et augmentant ainsi la confiance dans la prise de décision.

L'unique différence entre un système CADx et un système CBMIIR réside dans la représentation du diagnostic. En effet, l'ultime but consiste à renforcer la performance du diagnostic assisté par ordinateur (CADx) basé sur un résultat numérique généré par un classifieur et dont les radiologues ne sont pas familiarisés.

Dans les systèmes d'aide au diagnostic médical (CADx), un classifieur est optimisé sur un ensemble d'apprentissage, ensuite; la classification d'un nouveau cas d'entrée sera réalisée.

Evidemment, les radiologues reçoivent l'aide à travers des symboles numériques indiquant par exemple la classe dont appartient la région d'intérêt requête.

Les radiologues se plaignent du fait que la sortie abstraite d'un classifieur n'est pas évidente. En effet, ils se sentent plus confiant lors de la réalisation de leur diagnostic final lorsque des cas déjà diagnostiqués et similaires, au cas en cours d'analyse, sont récupérés et affichés à partir de bases de données indexées de cas déjà analysés ainsi que le résultat de la classe à qui appartient le cas en cours d'analyse. Entre autre, les radiologues peuvent indexer de nouveaux cas pathologiques.

Rappelons que les radiologues acquièrent leur formation en radiologie par des mammographies de lecture, leur expérience est acquise par la lecture d'un grand nombre de cas, qui les parviennent à reconnaître l'anatomie normale, l'anomalie bénigne, et en quoi consiste le cancer. Pour approfondir les connaissances dans ce sens, l'auteur est dirigé vers une étude sur l'État de l'art des approches d'indexation et de recherche des mammographies par le contenu (CBMIIR) [37]

III.7. Approche proposée basée fusion des lois de puissance : Zipf, Zipf inverse et les SFTA pour la classification des zones d'intérêt des mammographies

Comme il a été mentionné dans le premier chapitre, de nombreuses méthodes d'analyse de texture existent, et peuvent être appliquées pour l'extraction des descripteurs relatifs aux mammographies numériques.

Nous contribuons une fusion entre les descripteurs texturaux générés par les lois de Zipf et de Zipf inverse avec une technique d'analyse de texture, puissante et récente, connue sous le nom SFTA (Segmentation-based Fractal Texture Analysis) pour aboutir à une analyse judicieuse de la texture des mammographies par l'apport complémentaire de ces deux techniques d'analyse texturale.

III.7.1. 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 permettent de quantifier la structure sous-jacente du contenu de l'image pour analyser et caractériser la complexité structurelle de la texture, elles sont basées sur une approche statistique. Pour extraire les descripteurs de la texture au sein de l'image mammaire, une technique de segmentation se basant sur les lois de Zipf et de Zipf inverse a été proposée et présentée dans [49] afin d'explorer uniquement les zones d'intérêt des mammographies. Une fois que nous extrayons ces zones, un pré-traitement réalisé avec le codage des rangs généraux (mentionné dans la sous section II.4.2) est appliqué sur les pixels des zones d'intérêt, par la suite nous analysons les zones d'intérêt avec les lois de Zipf et Zipf inverse pour obtenir les courbes de Zipf et Zipf inverse, comme montré sur la figure (III.5) [50]. Où cette dernière présente 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. Notons que dans le cas des images mammaires seines nous avons utilisé un découpage d'une partie du tissu sein.

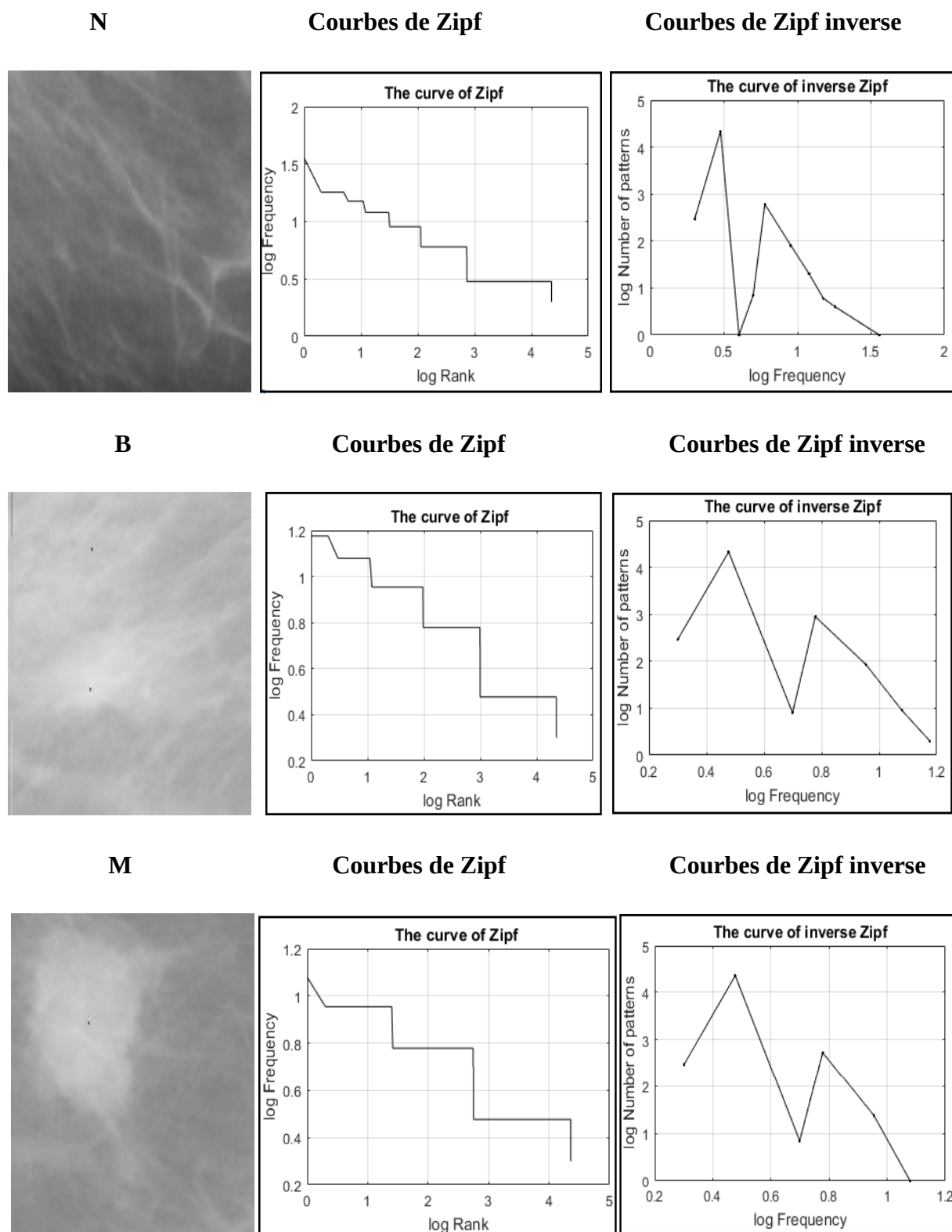


Figure III.5 Courbes de Zipf et de Zipf inverse de trois zones d'intérêts encodées par le codage des rangs généraux : **N** zone d'intérêt sans tumeur, **B** zone d'intérêt portant une tumeur bénigne, et **M** zone d'intérêt portant une tumeur maligne.

- **Analyse des courbes de Zipf relatives aux diverses zones d'intérêt**

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

Tandis que pour le cas de la zone d'intérêt présentant une tumeur bénigne nous observons une ordonnée à l'origine plus basse. Alors que pour la zone d'intérêt présentant une tumeur maligne, nous observons une ordonnée à l'origine de la courbe de Zipf nettement plus basse que celles de la tumeur bénigne du fait que la tumeur maligne présente une texture hétérogène causée par la nature invasive des tumeurs malignes et par conséquent 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 aux diverses zones d'intérêt**

Nous pouvons attester que la distinction des courbes de Zipf inverse 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 est claire, 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, en effet, 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 [37].

De cela, nous pouvons dire que l'analyse des zones d'intérêts avec les lois de Zipf et de Zipf inverse en utilisant un pré-traitement basé sur le codage des images mammaires par le codage des rangs généraux peut effectuer de manière significative la distinction entre les tumeurs malignes ou bénignes ainsi que le tissu sein à travers les descripteurs calculés à partir des courbes de Zipf et de Zipf inverse de chaque zone d'intérêt, les descripteurs que nous extrayons à partir des courbes de Zipf et de Zipf inverse sont les suivants [50]:

- ❖ **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 III.1 :

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

❖ **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 III.2 :

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

❖ **Entropie1 de la courbe de Zipf**

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

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

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:

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

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**

❖ **La constante 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.

III.7.2. Analyse et caractérisation de la texture des zones d'intérêt par les SFTA

Pour analyser la texture du tissu mammaire avec les SFTA (Segmentation-based Fractal Texture Analysis), les descripteurs sont extraits à partir des zones d'intérêt (ROI : Region Of Interest) segmentées sur les mammographies en les décomposant en un ensemble d'images binaires à partir duquel les dimensions fractales sont calculées [51]. Ces dimensions sont utilisées pour décrire les motifs de texture segmentés des mammographies. Pour la décomposition des zones d'intérêt (ROIs) d'échelle de gris en entrée, la méthode de Two-Threshold Binary Decomposition (TTBD) est appliquée. Dans ce type de décomposition, un ensemble de nt nombre de valeurs de seuil de région d'intérêt en échelle de gris est calculé en appliquant récursivement l'algorithme Ostu multi niveau, où nt est le paramètre défini par l'utilisateur (nous avons procédé par expérimentation et $nt=4$ nous a semblé la valeur optimale, en effet, lorsque la valeur de nt est plus grande, le nombre des descripteurs redondants augmente et par conséquent la complexité et les besoins en mémoire augmentent aussi). Un ensemble de zones d'intérêt binaires (bROI) est calculé à partir d'une ROI à échelle de gris en appliquant une Two-Threshold Binary Decomposition (TTBD). Dans cette méthode, des paires de valeurs inférieure et supérieure de ce seuil (t_l, t_u) sont sélectionnées parmi l'ensemble des valeurs de seuil, T et le ROI d'échelle de gris est décomposé comme [52]:

$$bROI(x, y) = \begin{cases} 1 \\ 0 \end{cases}, \text{ if } t_l < ROI(x, y) \leq t_u \quad (III.5)$$

Dans une Two- Threshold Binary Decomposition (TTBD) de ROI à échelle de gris, un nombre $2nt$ de bROI est obtenu en utilisant toutes les paires de valeurs de seuil de l'ensemble T , où nt est le nombre de valeurs de seuil. Après la décomposition de l'image binaire, trois composantes telles que la dimension fractale, le niveau de gris moyen, et la zone de région des bROI résultantes sont calculées. La dimension fractale de bROI est générée à partir d'une texture de son image de bordure. L'image de bordure est la région des limites de bROI calculée en utilisant l'ensemble des 8 pixels connectés. La frontière d'image a la valeur un au pixel (x, y) sur le bROI $(x, y) = 1$ correspondant, et ayant au moins un pixel voisin zéro. Sinon, l'image de bordure prend la valeur zéro. La dimension fractale (D) mesure le degré d'irrégularité d'une image. Différentes approches sont là pour calculer la dimension fractale d'une image, mais le comptage en boîte est la méthode la plus courante à cet égard. La dimension fractale (D) d'un bROI est calculée à l'aide d'une grille de carrés ayant chacun la taille s et donnée comme [52]:

$$D = - \lim_{s \rightarrow 0} \frac{\log(N_s)}{\log(s)} \quad (\text{III.6})$$

Où, N_s est le nombre de carrés nécessaires pour couvrir la portion de bROI. Maintenant, le vecteur de caractéristiques est construit pour chaque région d'intérêt en utilisant la collection en série des trois composantes, la dimension fractale, le niveau de gris moyen et la zone de région des bROI résultantes. La taille du vecteur de caractéristiques résultant est 1×21 . Chaque image donne 8 images binaires, car $nt = 4$.

III.7.3. Bilan

Notre approche consiste donc en la caractérisation de la texture des régions d'intérêt par une fusion de deux méthodes : les lois puissance Zipf et Zipf inverse ainsi que les SFTA dans le but de bénéficier de l'apport complémentaire de ces deux approches d'analyse texturale.

L'ultime but demeure l'obtention d'une classification pathologique des zones d'intérêt, la plus précise possible, pour donner un deuxième avis aux radiologues durant la réalisation du diagnostic.

III.8. Conclusion

Le cancer du sein est le type de cancer le plus courant chez les femmes dans le monde. La mammographie est considérée comme l'outil le plus efficace pour la détection et le diagnostic. Cependant, l'interprétation de la mammographie est une tâche répétitive et donc sujette aux erreurs. Caractérisée par au moins deux tâches difficiles à savoir : la recherche des mammographies et le diagnostic des cas représentés par les mammographies.

Une telle situation a nécessairement engendré chez les professionnels le besoin d'avoir des systèmes l'indexation et de recherche des mammographies par le contenu (CBMIIR) assisté par ordinateur assurant ces deux tâches très importantes, comme expliqué dans ce chapitre.

Dans ce chapitre, nous avons présenté l'approche contributive pour la caractérisation de la texture des zones d'intérêt extraites des mammographies. En effet, nous nous basons sur cette dernière pour le développement de notre système d'aide au diagnostic médical du cancer du sein assisté par ordinateur. Le prochain chapitre fera l'objet de davantage de détails dans ce sens.

CHAPITRE IV

IV.1. Introduction

L'indexation des images médicale par le contenu désigne un domaine ambitieux qui est rendu possible par les matériels médicaux actuels ainsi que les progrès scientifiques du traitement d'images et de la vision par ordinateur.

L'objectif de notre travail est de développer un système réalisant la classification des tumeurs du sein se basant sur 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 SFTA.

Dans ce sens, ce chapitre sera consacré au développement de notre système d'indexation et de recherche des mammographies par le contenu « Mammo_ZS» pour la classification des tumeurs du sein afin d'en tirer une variante dans le but d'offrir un deuxième avis au radiologue durant la réalisation du diagnostic du cancer du sein.

IV.2. Problématique

Notre problématique de recherche consiste à proposer une technique puissante d'analyse de la texture des mammographies pour l'étendre vers la vision par ordinateur. En effet, généralement les relations linéaires comme un polynôme ou une fraction rationnelle sont le plus souvent utilisés pour mettre en évidence le contenu d'une image.

Par contre, ces dernières sont limitées vu qu'il est difficile de modéliser, avec précision, un phénomène complexe tel que l'image au moyen de simples modèles linéaires.

Notons que les modèles les plus complexes sont rarement exploités, nous pouvons citer : les lois gaussiennes, les lois exponentielles et les lois puissance.

Dans [37] et [50], 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 que ces lois sont incontournables dans ce domaine. En effet, l'étude de la texture au sein des mammographies par ces lois permet de générer des descripteurs texturaux qui sont importants pour le diagnostic, mais ne sont pas facilement extraits visuellement par les radiologues.

Ainsi, les descripteurs de texture sont susceptibles de capter des caractéristiques qui sont importantes pour le diagnostic, mais difficilement extraites visuellement.

L'ultime but de notre travail est de travailler sur les perspectives de ces travaux en proposant une fusion des lois de Zipf et de Zipf inverse avec l'approche des SFTA dans le but 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.

Les images médicales jouent un rôle primordial dans le diagnostic des patients, et de nos jours, un nombre élevé de mammographies est constaté par jour, à cet effet, l'exploitation efficace de ces ressources demeure un défi pour les experts du domaine. La nécessité de concevoir des systèmes précis qui permettent un stockage et une indexation efficaces de ces images est l'une des tâches les plus importantes. En effet, les radiologues, à qui les systèmes d'aide au diagnostic médical du cancer du sein sont destinés, semblent insatisfaits 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. Dans ce sens, nous visons à 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), en effet, les radiologues se sentent plus sûres et confiants dans leur diagnostic lorsque des cas déjà diagnostiqués et similaires, à celui en cours d'analyse, sont récupérés et affichés à partir de bases de données indexées de cas déjà analysés ainsi que le résultat de la classe à qui appartient le cas en cours d'analyse.

Donc nous allons proposer un système d'indexation et de recherche des images mammaires se basant sur la fusion des lois de Zipf et Zipf inverse avec les SFTA pour la caractérisation de la texture des zones d'intérêt extraites à partir des mammographies pour ensuite réaliser un diagnostic assisté par ordinateur se basant sur la technique du template-matching.

IV.3. Classification

La classification est généralement la dernière étape dans le développement d'un algorithme de vision par ordinateur, qui consiste à prendre les caractéristiques extraites de l'image et de les utiliser pour classifier automatiquement les objets d'image.

La notion de classification signifie l'affectation d'une étiquette à des échantillons d'une base de données par l'utilisation d'un certain nombre de descripteurs. Ces descripteurs doivent être discriminants pour l'identification de chaque échantillon. Notons qu'en traitement d'image et vision par ordinateur, l'échantillon peut désigner un pixel, une zone dans l'image, un objet représenté dans l'image ou l'image entière. La classification a bien évidemment un but qui est soit de [53]:

- Classifier les pixels de l'image en différentes zones. Par conséquent, le problème de classification devient un problème de segmentation d'image en différents objets. Citons l'exemple de classifier les différentes zones d'une image mammaire en lésion ou non lésion.
- Classifier l'image ou les objets de l'image qui se trouvent dans les images. Citons l'exemple de classifier les zones d'intérêt segmentées dans les mammographies en malignes ou bénignes.

Nous distinguons deux catégories de méthodes de classification [53]:

IV.3.1. Les méthodes de classification non-supervisées

La classification non-supervisées, dite automatique, ou groupement connue en anglais par (clustering) consiste à déterminer les différentes classes naturellement sans aucune connaissance préalable. L'objectif, dans ce cas, est d'identifier une structure dans les images de la base en se basant sur leurs contenus. Les images sont attribuées aux différentes classes estimées selon deux critères essentiels qui sont la grande homogénéité de chaque classe et la bonne séparation entre les classes. Les méthodes les plus réputées: l'algorithme des k-moyennes, le regroupement hiérarchique.

IV.3.2. Les méthodes de classification supervisées

Si l'utilisateur possède suffisamment d'informations sur la population à étudier (tel est le cas des images mammaires), il peut effectuer une classification supervisée. Cette catégorie suppose avoir un groupe d'individus de chaque classe, dont on connaît leur appartenance.

Ces individus forment des échantillons "d'apprentissage". Ils sont utilisés pour entraîner le classifieur. D'autres échantillons, dits "de test", servent à valider la classification en évaluant sa pertinence à travers le taux d'individus bien classés. Il existe plusieurs méthodes de classification supervisées. Les méthodes les plus réputées: Machine à Vecteurs de Support (SVM), Réseau de Neurones, Méthode des K Plus Proches Voisins (KPPV), Arbre de Décision.

Dans notre travail, nous utilisons les méthodes de classification supervisées basée template-matching à travers les K Plus Proches Voisins (KPPV) ainsi que les Machines à Vecteurs de Support (SVM).

IV.4. Classification basée Template-matching

C'est une technique utilisée pour catégoriser les objets. Le but est de trouver des occurrences d'un modèle, qui est une sous-image dans une image plus grande. Les techniques Template matching comparent des parties d'images les unes par rapport aux autres. L'image échantillon peut être utilisée pour reconnaître des objets similaires dans l'image source. Template matching a diverses applications et est utilisée dans des domaines tels que la reconnaissance faciale (système de reconnaissance faciale) et le traitement d'images médicales.

Notre méthodologie pour la classification des zones d'intérêts segmentées au sein des mammographies à des zones d'intérêts portant des tumeurs bénignes, des tumeurs malignes ou des zones d'intérêts normales sans lésions se base sur l'approche template-matching, nous utilisons une technique de recherche pour trouver dans la base de données, les zones d'intérêts qui sont pathologiquement similaires à chaque zone d'intérêt requête. Par la suite, nous assignons à la zone d'intérêt requête sa classe pathologique selon le résultat donné par l'algorithme des K plus proches voisins (KPPV) ou celui des SVM, en effet, la classe pathologique majoritairement renvoyé par l'un de ces deux algorithmes sera la classe pathologique de la zone d'intérêt en cours de diagnostic.

IV.5. Conception

IV.5.1. Mammo_ZS : le système proposé pour l'indexation et la recherche des mammographies par le contenu

Notre système d'indexation et de recherche des mammographies par le contenu s'appuie sur l'objectif principal de prise de décision dans le contexte d'images similaires récupérées appartenant à des cas stockés dans une base de données de référence, fournissant ainsi une aide visuelle aux radiologues.

IV.5.1.1. Les Etapes suivies et accomplies pour la réalisation du système «Mammo_ZS»

1. Application des lois puissance de Zipf et de Zipf inverse et des techniques des SFTA (Segmentation-based Fractal Texture Analysis) pour l'analyse de la texture

Nous allons utiliser simultanément des descripteurs obtenus suite à l'analyse des zones d'intérêt par les lois puissance Zipf et Zipf inverse ainsi que ceux obtenus à partir des SFTA, en effet, nous désirons bénéficier de l'apport complémentaire de ces deux méthodes d'analyse de la texture (Comme mentionné dans section III.7).

2. Réduction de la dimension du vecteur descripteur par l'analyse en composantes principales

- **L'objectif de l'utilisation de l'ACP**

Nous allons appliquer l'analyse en composantes principales (ACP) sur les vecteurs descripteurs des zones d'intérêt (8 descripteurs générés par les lois de Zipf et de Zipf inverse et 8 autres générés par les SFTA) pour :

- Comprimer la taille du vecteur descripteur en ne conservant que l'information importante pour réduire le nombre de descripteurs à utiliser.
- L'obtention de 10 descripteurs non corrélés consistant en des combinaisons linéaires des descripteurs de départ, par la suite, nous utiliserons ces descripteurs dans les méthodes de modélisation, plus précisément, l'analyse discriminante afin de classer une zone d'intérêt requête en tumeur maligne ou bénigne ou bien une région se ne présentant pas de lésion.

L'analyse en composantes principales (ACP) est une procédure statistique qui souvent utilisée pour omettre les cotes corrélées en transformant les dimensions originales en de nouvelles dimensions qui sont une somme linéaire des dimensions originales mais qui sont linéairement décorréliées, appelées composantes principales. Ensuite, ces nouvelles dimensions sont classées en fonction de la variance, c'est-à-dire la dimension qui explique le plus de variabilité dans les données obtient le premier rang et ainsi de suite [54].

L'ACP est effectué par décomposition de la valeur propre de la matrice de corrélation de données après la normalisation des données pour chaque dimension. ACP fournit des scores et des chargements. Les scores sont les valeurs transformées correspondant à chaque point de données et les chargements sont les coefficients avec lesquels la variable d'origine doit être multipliée pour obtenir le score.

Supposons que les données soient sous la forme d'une matrice $m \times n$ X où m est le nombre d'observations et n est leur dimensionnalité. La moyenne est calculée le long de chacune des n dimensions ($i = 1 \dots n$) produisant un vecteur ligne de moyennes u .

$$u[i] = \frac{1}{m} \sum_{j=1}^m X[j, i] \quad (\text{IV.1})$$

Les moyennes sont soustraites de chaque ligne de la matrice d'observation X pour produire une matrice X_o . La matrice de covariance C est calculée en utilisant l'équation:

$$C = \frac{1}{m-1} X_o^T * X_o \quad (IV.2)$$

En trouvant les vecteurs propres de la matrice de covariance C , ils sont disposés dans l'ordre ascendant des valeurs propres correspondantes. Les premiers vecteurs propres forment l'ensemble des vecteurs de base sur lesquels les observations originales sont projetées. La valeur K est habituellement choisie de sorte que la variance dans les données projetées ne soit pas inférieure à une certaine fraction de la variance des données d'origine. La figure IV.1 résume les deux étapes précédentes:

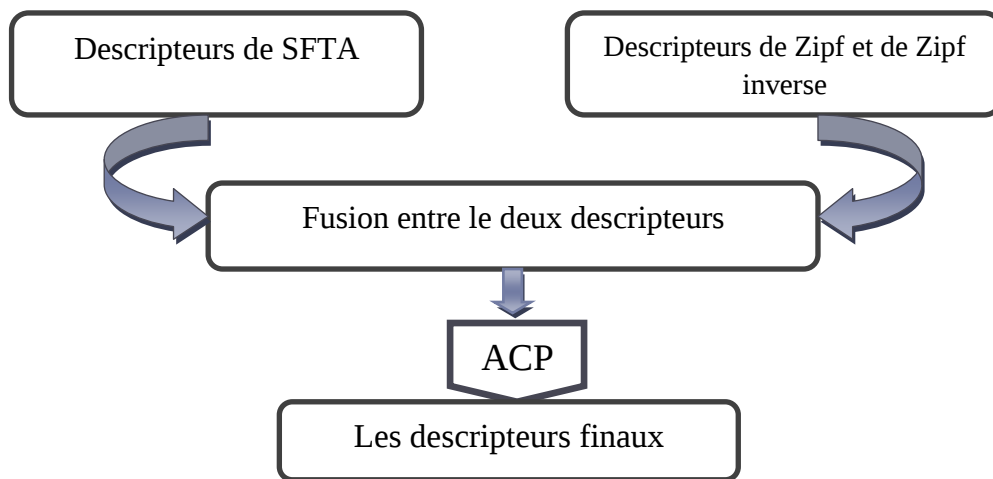


Figure IV.1 Le processus d'obtention des descripteurs finaux.

3. Template-matching et classification (diagnostic)

L'étape de template-matching vise à trier tous les descripteurs des zones d'intérêt indexées en fonction de leur similarité avec le vecteur descripteur de la requête. Ceci est obtenu en calculant la distance euclidienne d entre deux vecteurs caractéristiques, sur laquelle se base l'algorithme des KPPV.

L'algorithme *k-plus proche voisin (kppv)* est une méthode supervisée de classification classique. Pour classer une instance inconnue représentée par une fonctionnalité vecteurs comme un point dans l'espace des caractéristiques, l'algorithme *kppv* calcule les distances entre le point et les points de l'ensemble de données d'apprentissage. Habituellement, la distance euclidienne est utilisée comme mesure de similarité (ou fonction de distance). Ensuite, il assigne le point à la classe parmi ses k plus proches voisins (où k est un entier).

Fonction de distance

Pour mesurer la distance entre les points A et B dans un espace de caractéristiques, différentes distances fonctions 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 généralement utilisée par [55]

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^m (x_i - y_i)^2}{m}} \quad (IV.3)$$

Pour la classification, nous utiliserons aussi les machines à vecteurs de support (SVM).

Les machines à vecteurs de support (SVM) appartiennent à une méthode supervisée de classification. L'idée de base derrière l'algorithme SVM est 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.

Si les points ne sont pas linéairement séparables, les points sont projetés dans un nouvel espace en utilisant une fonction du noyau (kernel).

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.

SVM multiclasse, L'approche dominante le cas multiclasse est de réduire le problème multiclasse unique en plusieurs problèmes de classification binaires. Les méthodes communes pour une telle réduction incluent [56]:

Construire des classificateurs binaires qui distinguent (i) entre l'un des labels et le reste (un-contre-tous) ou (ii) entre chaque paire de classes (un-contre-un,). La classification des nouvelles instances pour le cas un-contre-tout est faite par une stratégie gagnant-preneur, dans laquelle le classificateur avec la fonction de sortie la plus élevée affecte la classe (il est important que les fonctions de sortie soient étalonnées pour produire des scores comparables) . Pour l'approche un-contre-un,, la classification est faite par une stratégie de vote max-wins, dans laquelle chaque classificateur assigne l'instance à l'une des deux classes, puis le vote pour

la classe assignée est augmenté d'un vote, et finalement la classe avec le plus de votes détermine la classification de l'instance.

IV.5.1.2. L'architecture globale du système «Mammo_ZS»

La figure IV.2 présente l'architecture de notre système Mammo_ZS.

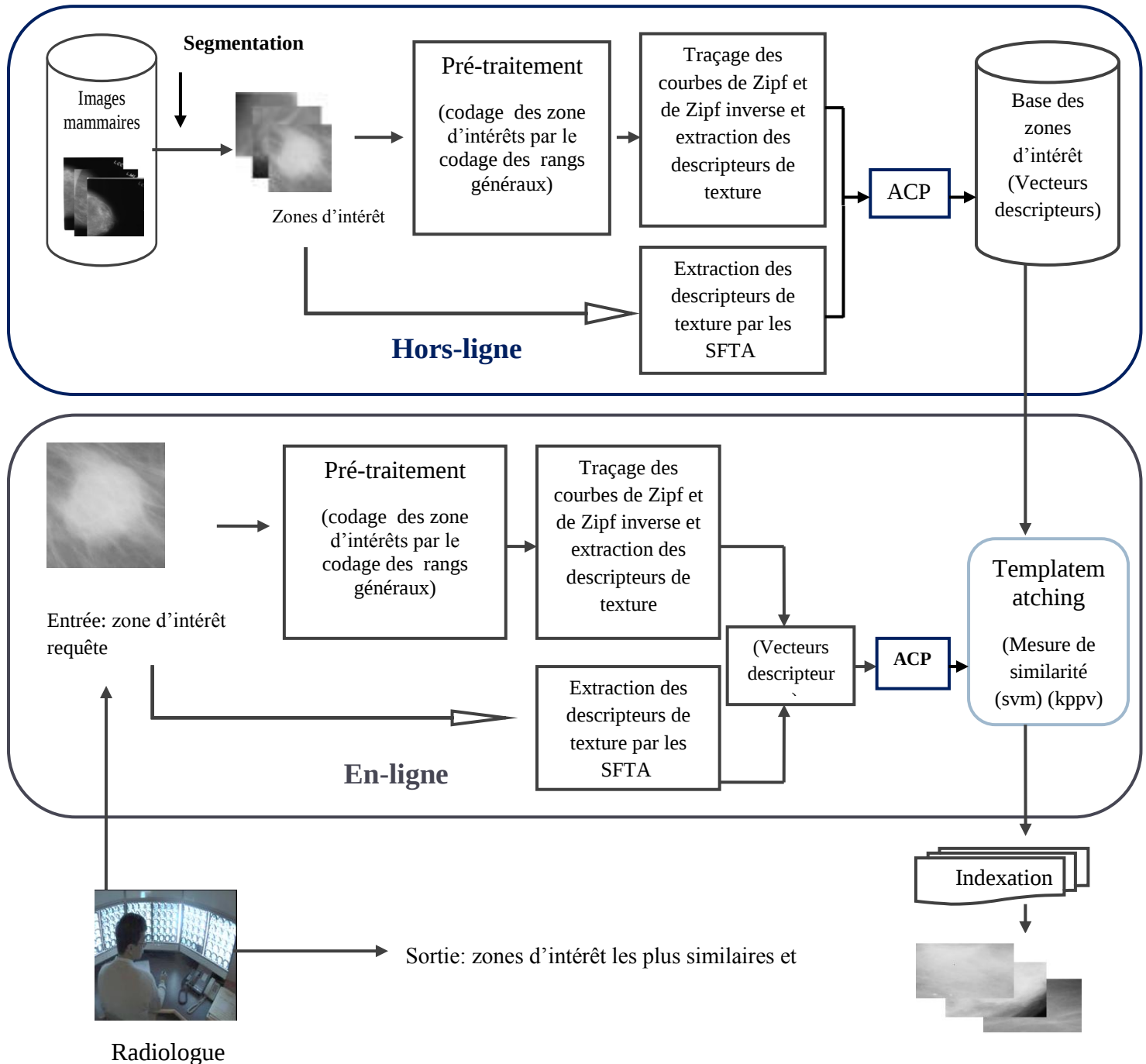


Figure IV.2 Architecture global du système Mammo_ZS proposé.

Nous détaillons les deux phases de notre système proposé comme suit :

La phase hors ligne :

- L'analyse des zones d'intérêt par les lois de Zipf et de Zipf inverse dans le but de tracer les courbes de Zipf et Zipf inverse et extraire les 8 descripteurs texturaux générés par ces lois puissance.
- Extraire les 8 descripteurs texturaux par SFTA.
- Diminuer la dimension du vecteur descripteur à 10 descripteurs suite à l'analyse en composantes principales (ACP).
- Stocker les vecteurs descripteurs dans la base de données.

La phase en ligne :

- Dans cette phase, le radiologue intègre une zones d'intérêt requête, segmentée à partir d'une mammographie ; au système pour identifier sa pathologie.
- Pour chaque zone d'intérêt segmentée à partir d'une mammographie, un vecteur de descripteurs texturaux est calculé suite à son analyse par les lois de Zipf et de Zipf inverse.
- Extraire les descripteurs texturaux par SFTA à partir de cette zone d'intérêt.
- Appliquer l'analyse en composante principale (ACP) sur le vecteur descripteur de la zone d'intérêt requête pour générer les nouveaux descripteurs.
- La mesure de similarité (matching) entre le vecteur descripteur de la région d'intérêt requête et le vecteur descripteur de chaque région d'intérêt dans la base de données est mesuré par la distance Euclidienne intégrée dans l'algorithme des K plus proches voisin.
- Les cas récupérés par le système suite à la recherche des K plus proches voisin de la région d'intérêt requête s'utilisent comme prédicteurs utiles de la classe pathologique de la requête courante. Ceci, en attribuant à la zone d'intérêt requête la classe pathologique majoritairement renvoyée par le processus de recherche.
- Les machines à vecteurs de support (SVM) ont aussi été utilisées pour le processus de classification.

IV.6. Implémentation

IV.6.1. L'environnement de développement

Notre application a été réalisée sur un PC de type ASUS ZENBOOK 8.00 GB de RAM sous Windows 8.1 Pro.

Les langages utilisés sont : Matlab version 2017a.

MATLAB (matrix laboratory) est un environnement informatique numérique multi-paradigme. 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, Java, et Python...

On a choisit 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.

IV.6.2. 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 [37]. Nous avons segmenté des régions d'intérêts de 155x155 pixels.

IV.6.3. 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 (80 images mammaires) et nous avons consacré le reste de l'échantillons pour le test.

La performance du système proposé a été évaluée avec l'aide de paramètres, notamment la précision de la classification (%) et le temps pris avec et sans l'analyse en composante principale (ACP).

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

Table IV.1 Variation de la précision de classification du système «Mammo_ZS» pour diverses valeurs des k plus proche voisins.

Les valeurs du K	Précision de la classification (%)
3	50%
4	57.5%
5	62.5%
8	60%
10	57%
15	55%

Nous pouvons observer que la précision de la classification varie entre 50% et 62.5%, en effet, lorsque k augmente, la précision du système MammoZS diminue de façon constante pour des valeurs du K comprises entre 8 et 15. La meilleure performance du système proposé a été obtenue pour $k=5$. En effet, en prenant un nombre considérable des k plus proches voisins, ceci génère l'apparition du bruit dans les cas d'apprentissage qui sont utilisés pour la décision de la classification du cas en cours d'analyse.

Le tableau IV.2 évoque les résultats de la précision de classification et le temps requis pour le système «Mammo_ZS» en utilisant les K -PPV et les SVM, avec et sans l'analyse en composante principale (ACP):

Table IV.2 Performance du système «Mammo_ZS» avec et sans ACP.

les méthodes les paramètres	Avec ACP		Sans ACP	
	K-PPV	SVM	K-PPV	SVM
Précision de classification(%)	62.5%	60%	60%	35%
Temps requis	0.093750000000 000	2.812500000000 00	0.390625000000 00	3.546875000000 0

Les résultats montrent que la performance du système «Mammo_ZS» est meilleure lorsque nous appliquons l'ACP pour la réduction du volume des vecteurs descripteurs. Entre autre, l'algorithme des SVM aboutit à une précision de classification (60%) proche de l'algorithme KPPV (62.5%) lorsque une analyse en composante principale est appliquée.

IV.6.4. Présentation de quelques vues du système «Mammo_ZS»

Le système «Mammo_ZS» est doté d'une fenêtre login permettant d'accéder aux différentes fonctionnalités du système. Ceci est mentionné sur la figure IV.3.

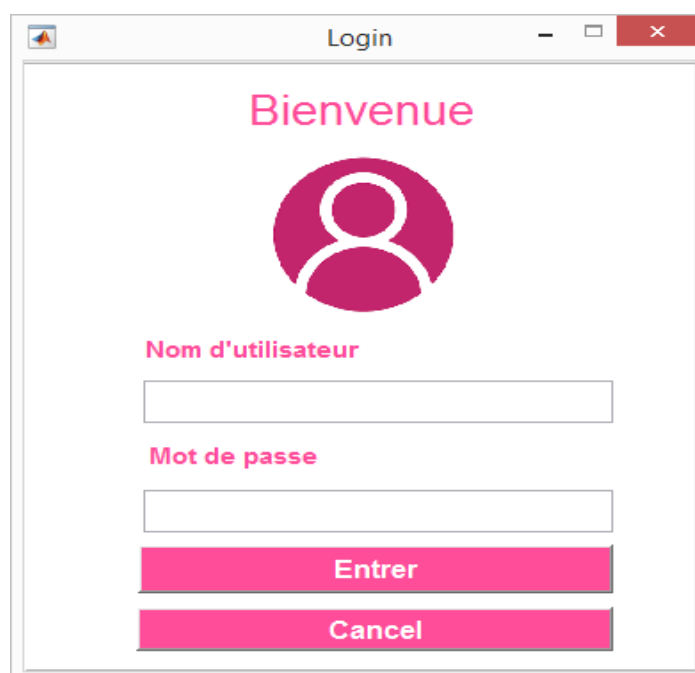


Figure IV.3 Fenêtre Login du système «Mammo_ZS».

En exécutant le bouton «Entrer», nous accédons à l'interface principale du système «Mammo_ZS» après avoir introduit le nom d'utilisateur et le mot de passe bien évidemment comme mentionné sur la figure IV.4..



Figure IV.4 Interface principale du système «Mammo_ZS».

En exécutant le bouton «À propos», nous aurons des détails sur le travail proposé comme exposé sur la figure IV.5.



Figure IV.5 Fenêtre des détails de réalisation de ce travail.

En exécutant le bouton «Entrer» de la fenêtre principale, la fenêtre de la figure IV.6 apparaît. En cliquant sur le bouton « Parcourir », le radiologue choisira la zone d'intérêt requête qu'il désire diagnostiquer.

Sur cette figure, nous démontrons également la sortie du système Mammo_ZS 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 normal. Comme mentionné sur cette figure, 3 régions d'intérêt parmi les 5 renvoyées par le processus de recherche portent une pathologie bénigne, aidant le radiologue à diagnostiquer la région d'intérêt requête comme tumeur bénigne ce qui est vraiment le cas. En outre, le système Mammo_ZS offre un affichage des voisins appartenant à la même classe pathologique de la région d'intérêt requête.

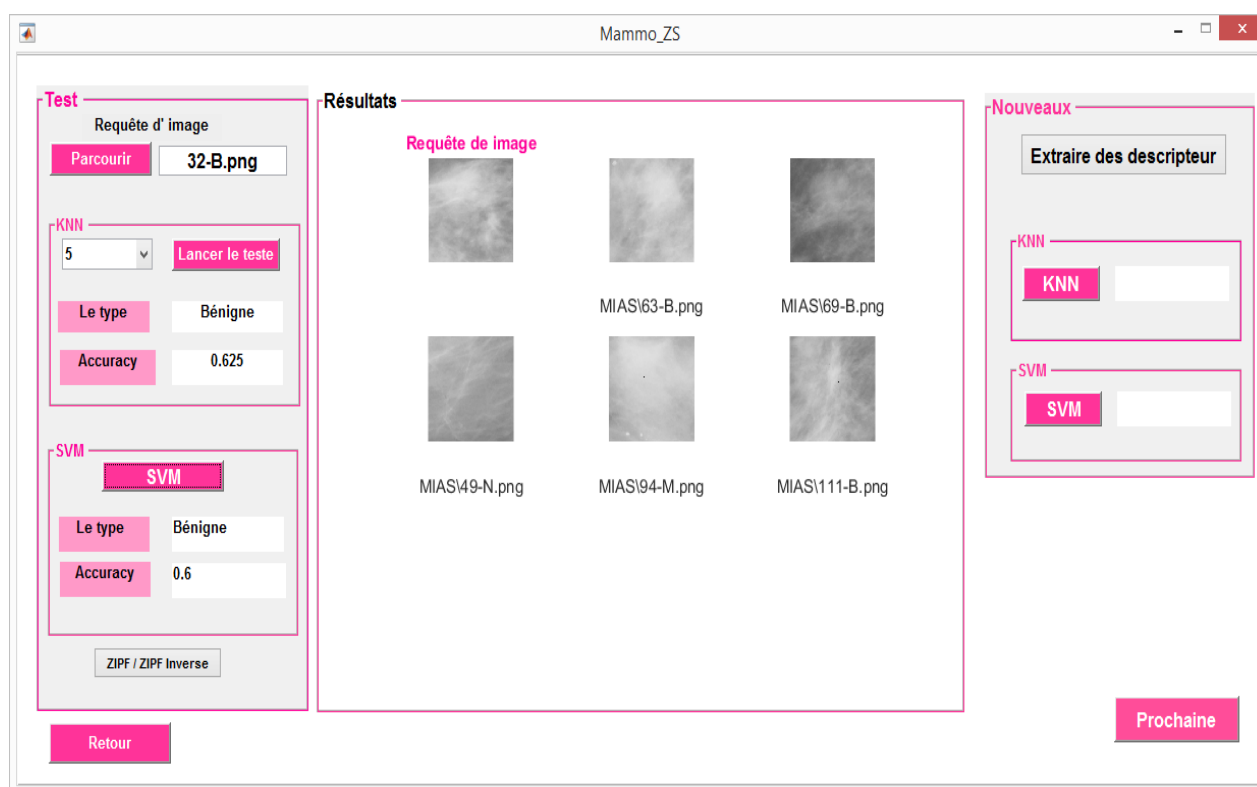


Figure IV.6 Processus de décision médicale issue de l'utilisation de la sortie de Mammo_ZS par la considération de $k=5$.

La région «Nouveaux» nous permet de traiter une nouvelle zone d'intérêt requête que pourra le radiologue introduire au système.

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

Table IV.3 L'approche proposée vis-à-vis des approches

Approche	Technique	Base de données	Précision de classification(%)
Approche proposée Zipf, Zipf inverse et les SFTA	KPPV	MIAS	62.5%
	SVM		60%
Zipf, Zipf inverse et les filtres de Gabor	KPPV		65%
	SVM		67.5%
Zipf, Zipf inverse et les Ondelettes	KPPV		80%
	SVM		70%
Zipf, Zipf inverse et les transformée de Fourier	KPPV		80%
	SVM		62%

IV.7.Conclusion

Dans ce chapitre, nous avons proposé un système d'indexation et de recherche de mammographies par le contenu «Mammo_ZS» 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 62.5%, ceci par les KPPV et de l'ordre de 60% par les SVM.

CONCLUSION GÉNÉRALE

L'image médicale comporte des informations importantes pour l'aide au diagnostic médical, ces dernière ont expansées de manière exponentielle en quantité, en raison de l'augmentation rapide du nombre de visites médicales par jour, ceci, a rendu de l'exploitation efficace des images médicales un défi pour les experts du domaine par la conception de systèmes précis qui permettent un stockage et une indexation efficaces de ces images.

Dans ce mémoire nous avons abordé le problème de la recherche d'images. Plus précisément, nous nous sommes focalisés à la recherche d'images basée sur le contenu, en appliquant simultanément deux techniques d'analyse de la texture aux régions d'intérêt, segmentées dans les mammographies, consistant en les lois puissance Zipf et Zipf inverse ainsi que les SFTA (Segmentation-based Fractal Texture Analysis). Dans le but de renforcer cette fusion de descripteurs texturaux, nous avons appliqué une analyse en composantes principales (ACP) qui est perçue à ce niveau comme une méthode de recherche de nouveaux descripteurs consistant en les composantes principales, non observables et non corrélés entre eux et qui désignent les descripteurs les plus influents et discriminants.

Notre système «Mammo_ZS» 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 précision de classification assez bonne de l'ordre de 62.5% par l'algorithme des K plus proches voisins, et une précision de classification de l'ordre de 60%) par l'algorithme des machines à vecteur de support.

Perspectives

Dans nos futurs travaux, nous cherchons à améliorer notre approche par l'étude de nouvelles techniques de caractérisation de la texture, de sélection et de classification des descripteurs.

Notre but ultime sera de générer une base de données volumineuse en recueillant des mammographies avec leurs étiquettes, de classes pathologiques, de différents centres de diagnostic afin que la validation des systèmes soit performante.

BIBLIOGRAPHIE

- [1] Szeliski, R. (2010). Computer Vision: Algorithms and Applications, Springer Science & Business Media
- [2] Borji A. (2017). Negative Results in Computer Vision: A Perspective. arXiv:1705.04402 [cs.CV]. page 1 - 9.
- [3] BOUCHER, A. (2013). Recalage et analyse d'un couple d'images: application aux mammographies. hèse de doctorat. Université Paris Descartes.
- [4] Dipert, B. Shoham, A.(2012). Eye, robot: Embedded vision, the next big thing in digital signal processing. IEEE Solid-State Circuits Magazine, pp. 26–29.
- [5] Schwambach, V. (2016). Methods and tools for rapid and efficient parallel implementation of computer vision algorithms on embedded multiprocessors. Thèse de doctorat. Université Grenoble Alpes.
- [6] Ayache, N. (2015). Des images médicales au patient numérique. Livre. Librairie Arthème Fayard et Collège de France.
- [7] Srinivasan, G. N. Shobha, G. (2008). Statistical Texture Analysis. WASET.ORG. pp.1264-1269.
- [8] Di Cataldo, S. Ficarra, E. (2017). Mining textural knowledge in biological images: Applications, methods and trends. Computational and Structural Biotechnology Journal. pp. 56–67.

- [9] Cheikhrouhou, I. (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein. Thèse de doctorat, Université d'Evry-Val d'Essonne.
- [10] Claude, I., Pouletaut, P., Huault, S., & Boulanger, J. C. (2001). Integrated color and texture tools for colposcopic image segmentation. International Conference on, pp. 311-314.
- [11] Haralick, R. M. (1979). Statistical and structural approaches to texture. In Proc. IEEE. pp. 786-804.
- [12] Haidekker, M.A. (2010). Texture Analysis. John Wiley & Sons, Inc., pp. 236–275.
- [13] Petrou, M. (2011). Texture in biomedical images. Springer-Verlag, Berlin, Heidelberg. pp. 157–176.
- [14] Linda G. Shapiro , George C. Stockman. (2001). Computer Vision, Upper Saddle River: Prentice–Hall.
- [15] Favorskayaa, M. Pyataevab, A. Popova, A.(2017). Texture analysis in watermarking paradigms. Procedia Computer Science. pp.1460–1469.
- [16] Haralick RM, Shanmugam K, Dinstein I.(1973). Textural features for image classification. IEEE. pp.610–621.
- [17] Machado,B. B. Junior, J. F. R. (2016). Texture analysis using complex system models: fractal dimension, swarm systems and non-linear diffusion. Biblioteca Digital da Produção Intelectual - BDPI. pp. 2409–2414.
- [18] Ribas, L. C. Gonçalves, W. N. Bruno, O. M.(2016). Texture analysis using deterministic partially self-avoiding walk with thresholds. WVC proceedings. pp. 39-44.
- [19] Hassaballah, M. Alshazly, H.A. Abdelmgeid, A.A.(2016). Image Features Detection, Description and Matching. Springer International Publishing Switzerland. pp. 11-45.
- [20] Dash, S. Senapati, M.R. Jena, U.R. (2018). K-NN based automated reasoning using bilateral filter based texture descriptor for computing texture classification. Egyptian Informatics Journa. pp. 1-12.
- [21] Danuser, G.(2011). Computer Vision in Cell Biology. Elsevier Inc. pp. 973-978.

- [22] Brian, L. DeCost, E.A.H. (2015). A computer vision approach for automated analysis and classification of microstructural image data. *Computational Materials Science*. pp. 126–133.
- [23] khodaskara ,A. Ladhakeb, S.(2015). Semantic Image Analysis for Intelligent Image Retrieval. *Procedia Computer Science*. pp.192 – 197.
- [24] Yaneer, B.Y. (2015). Concepts: Power Law. New England Complex Systems Institute.
- [25] Caron, Y. (2004). Contribution de la loi de Zipf à l'analyse d'images. Thèse de doctorat. Université de Tour.
- [26] Pareto, V. (1897). Cours d'Economie Politique, Vol. 2, F. Pichou, Paris.
- [27] Schaffar, A. (2009). Croissance et hiérarchie urbaines dans les pays émergents. Thèse de doctorat. Université de la Réunion, Université de Toulon Sud.
- [28] Pareto, V. (1895),(1965). La courbe de la répartition de la richesse, Librairie Droz, Genève.
- [29] Persky J., (1992). Pareto's law, *Journal of Economic Perspectives*. pp.181-192.
- [30] Auerbach, F.(1913). Das Gesetz der Bevölkerungskonzentration. reproduit dans *Regional Science and Urban Economics*. pp.601-615.
- [31] Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology. Addison-Wesley.
- [32] Yang, C. (2010). Who's Afraid of George Kingsley Zipf?. University of Pennsylvania.
- [33] Egmond, M. van.(2011). Word Finding Difficulties in Aphasia and their Effect on Zipf's Law. Master's thesis.
- [34] Baroni, M. (2009). Distributions in text. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics: An International Handbook* . pp. 803–822.
- [35] Dellandréa, E. l Makris,P. Vincent,N.(2004). Zipf Analysis of Audio Signals. World Scientific Publishing Company. pp. 73-85.
- [36] Pareti, R. vincent, N. (2006). Global method based on pattern occurrences for writer identification, IWFHR.

- [37] Hamoud, M. (2014). Indexation et segmentation d'images basées loi de Zipf et Zipf inverse. Thèse de doctorat. University Badji Mokhtar-Annaba.
- [38] Caron, Y. Makris, P. Vincent, N. (2007). Use of power law models in detecting region of interest, the journal of the Pattern Recognition, 40. pp. 2521–2529.
- [39] Laimeche, L. (2010). Détection des Informations Cachées dans les Images Numériques Basée loi de Zipf. Thèse de doctorat. Université de Tébessa (Cheikh Larbi Tbessi).
- [40] Caron, Y. Makris, P. Vincent, N.(2005). Caractérisation d'une région d'intérêt dans les images. Published in EGC.
- [41] Caron, Y. Charpentier, H. Makris, P. Vincent, N.(2003). Power Law Dependencies to Detect Regions of Interest. Springer-Verlag Berlin Heidelberg. pp. 495–503.
- [42] Carvalho, E. D. Carvalho Filho, A. O. Sousa, A. C. Gattass, M. (2018). Method of differentiation of benign and malignant masses in digital mammograms using texture analysis based on phylogenetic diversity. Computers and Electrical Engineering. pp 210–222.
- [43] Beura, S.(2016). Development of Features and Feature Reduction Techniques for Mammogram Classification. Thèse de doctorat. National Institute of Technology Rourkela.
- [45] Suzuki K. A. (2012). review of computer-aided diagnosis in thoracic and colonic imaging. Quantitative Imag Med Surg. pp.163–76.
- [46] Jalalian, A. Mashohor, SB. Mahmud, HR. Saripan, MI. Ramli, AR. Karasfi, B. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review. Clinical Imaging. pp 420–426.
- [47] Bozek, J. Mustra, M. Delac, K. Grgic, M. (2009). A Survey of Image Processing Algorithms in Digital Mammography. Advanced. in Multimedia. Signal. Processing. and Communication SCI, 231. pp. 631–657.
- [48] Charriere, K. (2015). Analyse et caractérisation temps réel de vidéos chirurgicales. Application à la chirurgie de la cataracte. Thèse de doctorat. Université de Bretagne Occidentale.

- [49] Hamoud, M. Merouani, H. F. (2011). Detection of a Region of Interest in the Images Based on Zipf Laws. In The Seventh IEEE. (SITIS'2011). Dijon, France. pp. 416 – 421.
- [50] Hamoud, M. Merouani, H. F. Laimeche, L. (2014). The power laws: Zipf and inverse Zipf for automated segmentation and classification of masses within mammograms. Springer-Verlag Berlin Heidelberg.
- [51] Costa, A. F. Humpire-Mamani, G. Traina, A.J. M. (2012). An efficient algorithm for fractal analysis of textures," in Proc. 25th Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, pp. 39-46.
- [52] Beura, S. Majhi, B. Dash, R. (2015). Automatic Characterization of Mammograms using Fractal Texture Analysis and Fast Correlation Based Filter Method. Association for Computing Machinery. ACM acknowledges. pp 85-91.
- [53] Cheikhrouhou, I. (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein. Thèse doctorat. Université d'Evry-Val d'Essonne.
- [54] Varun, J. (2015). Visual Observation of Human Emotions. Thèse de doctorat. Université Grenoble Alpes.
- [55] Li-Yu, H. Min-Wei, H. Shih-Wen, K. Chih-Fong, T. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus. pp 1–9.
- [56] Duan, Kai-Bo. Keerthi, S. Sathya (2005). Which Is the Best Multiclass SVM Method? An Empirical Study. Multiple Classifier Systems. Springer-Verlag Berlin Heidelberg. pp. 278–285.

Liens internet

- [44] www.cancer.ca/en/cancer-information/diagnosis-and-treatment/tests-and-procedures/mammography/?region=qc

Résumé

Les images médicales jouent un rôle essentiel dans le diagnostic des patients, de sorte que la détection et la classification de la tumeur est un endroit important et délicat pour le diagnostic précoce du cancer du sein.

Dans ce mémoire nous avons appliqué les lois de puissance Zipf et Zipf inverse en plus de la SFTA (Segmentation-based Fractal Texture Analysis), pour l'analyse des images mammaires pour l'extraction des descripteurs texturaux de chaque technique, puis fusionnez-les et appliquez la méthode l'Analyse Composante Principale (ACP). Nous avons utilisé la méthode Template-Matching basée sur la classification par la méthode des k-plus proches voisins (k-ppv) pour rechercher les images les plus similaires en mammographie à choisir et la classification de la classe pathologique de ces images (malignes, bénignes et normal) en utilisant des machines à vecteurs de support (SVM). Aider les radiologistes à prendre la décision finale.

Mots clés

Images médicales, Classification, lois de puissance Zipf et Zipf inverse, SFTA, ACP, Template-Matching, k-ppv, Mammographie, Classe pathologique, SVM.