

## Résumé

Cette thèse s'inscrit dans le cadre de la classification des tumeurs cérébrales en utilisant les forêts aléatoires testées sur une base des images IRM (BRATS, 2015). Dans un premier temps, nous allons effectuer une étape de caractérisation qui nous permettra d'extraire des attributs de textures, en utilisant deux méthodes qui sont reconnues dans le domaine de l'analyse de texture. Il s'agit de l'Analyse Fractale et Les Matrices de Cooccurrences, afin de construire des vecteurs caractéristiques pour ensuite créer des bases de données qui seront, par la suite, utiliser pour la classification, où on va utiliser une méthode qui s'appelle Les forêts aléatoires afin de détecter le type de la tumeur (Haut Grade (Maligne) ou bien Bas Grade (Bénigne)).

**Mots Clés** : Analyse de texture, Forêts Aléatoires, Caractérisation, Classification, Images IRM, Tumeurs cérébrales.

## Abstract

In this paper, we report a part of the classification of brain tumors using a base of MR scans (BRATS, 2015), at first, a characterization step is a priority that allows us, to extract texture features, using two methods that are recognized in the field of texture analysis. We talk about the Fractal Analysis and the co-occurrence matrices, due to construct feature vectors, then create databases which will be used for the classification, and for this latter we will use a method called « Random Forests », to detect the type of the tumor (High grade or Low grade).

**Keywords** : Texture analysis, Random Forests, characterization, Fractal Analysis, co-occurrence matrices, MR scans, brain tumors,

# REMERCIEMENTS

Mes cinq années d'études motivantes à Tlemcen, ont été très enrichissantes et très agréables. Ceci est dû à de nombreuses personnes que j'ai côtoyées une ou plusieurs années. Je voudrais donc les remercier ici.

Tout d'abord, je voudrais dire un énorme merci à mon directeur de mémoire, Mr « Mostafa El Habib Daho » Sa disponibilité, son aide précieux et sa patience qui ont été d'un secours remarquable afin de mener ce travail à bien. Merci pour cette énergie.

Je tiens à remercier les membres de mon jury pour l'accueil qu'ils ont su réserver à mon travail et pour les efforts consacrés à la lecture de ce manuscrit.

Je voudrais également témoigner ma reconnaissance aux enseignants de la faculté de technologie, et en particulier département de Génie Biomédicale.

Je tiens aussi à remercier Dr Bounouara AR pour son accueil, sa disponibilité, son écoute.

Un grand merci aussi à tous mes amis, leur soutien sans relâche, leur rigueur. Et en particulier Ahmed, Issam, Messouad, Khalil, Amine. Un merci spécial pour Chahrazed pour son aide précieux et j'espère que ceux que j'aurais oublié ne m'en tiendront pas rigueur.

Je remercie ma famille pour leur soutien, et particulièrement mes parents qui m'ont toujours laissé très libre dans mes choix et m'ont toujours encouragé dans ceux-là.

# *Dédicaces*

*Je dédie ce mémoire A*

*Mes Grands-parents,*

*Mes Chers parents,*

*Mes frères : Issam, Loffi et leurs femmes*

*Ma sœur Soumia et Son Mari Lamine*

*Et Mes Anges Mohammed Anes, Adam*

# Table des matières

Résumé .....	1
Liste de Figures.....	7
Liste de Tableau.....	8
Liste des Abréviations.....	8
Préface .....	9
Introduction générale .....	11
Problématique .....	12
Organisation du manuscrit .....	12
<b>CHAPITRE 1 CONTEXTE MEDICALE .....</b>	<b>14</b>
1.2 Un peu d'anatomie pour comprendre le cerveau .....	16
1.2.1 Que se passe-t-il à l'intérieur du cerveau ? .....	16
1.2.2 Les différentes parties du cerveau et leurs fonctions .....	17
1.2.3 Le cerveau, un organe sous haute protection .....	18
1.3 Les tumeurs cérébrales .....	19
1.3.1 Qu'est-ce qu'une tumeur ?.....	19
1.3.2 Tumeur de cerveau .....	20
1.3.3 Les causes des tumeurs de cerveau .....	21
1.3.4 Les symptômes d'une tumeur de cerveau .....	21
1.3.5 Gradation des tumeurs selon l'OMS .....	22
1.4 Diagnostic d'une tumeur de cerveau .....	23
1.4.1 Examen clinique et neurologique .....	23
1.4.2 Les examens d'imagerie.....	23
1.4.3 Analyser un échantillon de tumeur : L'examen anatomopathologique .....	23
1.5 Imagerie par résonance magnétique.....	24
1.5.1 Principe de fonctionnement .....	24
1.5.2 Avantages et inconvénients .....	25
1.6 Technologie et nouveaux traitements .....	25
1.7 Conclusion .....	27
<b>CHAPITRE 2 ETAT DE L'ART SUR LES METHODES D'EXTRACTION D'ATTRIBUT DE TEXTURE .....</b>	<b>28</b>
2.1 Introduction.....	29
2.2 La notion de texture.....	30

2.2.1	Définition .....	30
2.2.2	Type de textures .....	30
2.3	Etat de l'art des différentes méthodes d'extraction d'attribut de texture .....	31
2.3.1	Méthodes statistiques .....	31
2.3.1.1	Méthodes d'ordre un .....	31
2.3.1.2	Méthodes d'ordre deux .....	32
2.3.1.3	Méthodes d'ordre supérieur .....	32
2.3.2	Méthodes structurelles .....	33
2.3.3	Méthodes spatio-frequentielles.....	33
2.3.4	Méthodes basées sur un modèle.....	34
2.4	La classification .....	36
2.4.1	Définition :.....	36
2.4.2	Méthodes de classification :.....	36
2.5	Conclusion .....	38
<b>CHAPITRE 3 NOS METHODES DE CARACTERISATION ET CLASSIFICATION .....</b>		<b>39</b>
3.1	Introduction.....	40
3.2	Méthodes d'analyse de texture, application aux images IRM cérébrales .....	41
3.2.1	Méthodes basées sur les Matrices de Cooccurrences.....	41
3.2.1.1	Construction des matrices de cooccurrence .....	42
3.2.1.2	Les indices de Haralick .....	43
3.2.1.3	Choix de la méthode .....	44
3.2.2	Méthode d'Analyse Fractale .....	45
3.2.2.1	Self similarité ou auto similarité dans les images .....	46
3.2.2.2	Notion de la Dimension Fractale .....	46
3.2.2.3	Un algorithme efficace pour l'analyse fractale des textures : SFTA .....	47
3.2.2.4	Choix de la méthode « SFTA » .....	47
3.3	Classification avec les forêts aléatoires .....	48
3.3.1	L'apprentissage statistique .....	48
3.3.2	Les méthodes d'ensemble .....	48
3.3.2.1	Le Bagging .....	49
3.3.3	Classification and regression trees CART .....	53
3.3.4	Choix de la méthode .....	53
3.4	Conclusion .....	53

<b>CHAPITRE 4 CLASSIFICATION AVEC RANDOM FOREST, APPLICATION AUX IMAGES IRM CEREBRALE</b> .....	54
4.1 Introduction .....	55
4.2 Prétraitement d'images.....	55
4.2.1 Base de données BRATS, 2015.....	55
4.2.2 Conversion des images .....	56
4.3 Caractérisation et construction des bases de données.....	57
4.4 Résultat et Discussion .....	59
4.4.1 Pourquoi Nombre d'itération égale à 1 ? .....	59
4.4.2 Pourquoi Nombre d'arbre égal à 500 ? .....	60
4.4.3 Evaluations de la performance de classification : .....	61
4.4.4 Mesurer la performance de notre système .....	67
4.5 Application Java FX.....	69
4.6 CONCLUSION .....	72
CONCLUSION GENERALE.....	73
PERSPECTIVES.....	74
Bibliographie.....	75

## Liste de Figures

Figure 1.1 : Atlas de Vesale (image a) (source : De numani corporis fabrica), neurones colorés par la méthode de Golgi (image b) (source : Hubel) et phrénologie de F.J. Gall.

Figure 1.2 : les quatre lobes de l'hémisphère gauche

Figure 1.3 : la protection du cerveau

Figure 3.1 : objet Fractale

Figure 3.2 : principe du Bagging

Figure 3.3 : principe des Forêts Aléatoires

Figure 4.1 : interface graphique du logiciel *Slicer*

Figure 4.2 : interface graphique du logiciel *ImageJ*

Figure 4.3 : les différentes bases de données

Figure 4.4 : les différentes tailles des fenêtres des bases de données

Figure 4.5 : représentation du Taux de classification en fonction du nombre d'arbre

Figure 4.6 : représentation du Taux de classification en fonction du nombre d'arbre (2)

Figure 4.7 : présentation d'une partie de la base de données en utilisant la fenêtre  $16 \times 16$

Figure 4.8 : la différence entre les tailles des fenêtres

Figure 4.9 : Comparaison des meilleurs résultats obtenus

Figure 4.10 : présentation des classes du projet

Figure 4.11 présentation de l'interface principale de notre application

Figure 4.12 présentation de la fenêtre : Mode Window

Figure 4.13 présentation de la fenêtre : Result Window

## Liste de Tableau

Tableau 1.1 Gradation des tumeurs selon l’OMS

Tableau 4.1 Résultats de la classification en utilisant la fenêtre 16\* 16

Tableau 4.2 Résultats de la classification en utilisant la fenêtre 32\* 32

Tableau 4.3 Résultats de la classification en utilisant la fenêtre 64\* 64

Tableau 4.4 Résultats de la classification en utilisant la technique Analyse Fractale

Tableau 4.5 Résultats de la classification en utilisant la technique MC

Tableau 4.6 Résultats de la classification en utilisant la technique MC\*

Tableau 4.7 comparaison des résultats entre MC et MC\*

Tableau 4.8 Résultats de la classification en utilisant la combinaison {AF+MC}

Tableau 4.9 Résultats de la classification en utilisant la combinaison {AF+MC\*}

Tableau 4.10 La matrice de confusion de la technique RF (AF+MC)

## Liste des Abréviations

CAD : Computed Assisted Diagnosis

IRM : Imagerie par Résonance Magnétique

RMN : Résonance Magnétique Nucléaire

TEP : Tomographie par Emission de Positons

OMS : Organisation Mondiale de la Santé

DWT : DiscreteWavelet Transform

AF : Analyse Fractale

MC : Matrices de Cooccurrences

MC : Matrices de Cooccurrences avec seulement 6 attributs

BDD : Base de données

RF : Random Forest

CART : Classification And Regression Trees

SFTA : Segmentation-based Fractal Texture Analysis

HG : Haut grade

BG : Bas Grade

## Préface

Le Cerveau fascine, Il a pourtant été longtemps considéré comme un organe d'importance mineure dans le corps humain. Les égyptiens par exemple ne se préoccupaient pas de le conserver lors des momifications, car pour eux c'est le cœur qui est le centre de la pensée et des sentiments. Alcméon (500 av. JC) un médecin de Crotona en Grande Grèce du VIème siècle de notre ère, a été le premier occidental à avoir affirmé que « *ce qui gouverne siège dans le cerveau* » et Galien, le plus célèbre médecin de l'Antique gréco-latine, qui a vécu au IIème siècle, a forgé le terme d'*impression* pour exprimer l'idée que les souvenirs sont gravés dans le cerveau, qui n'était alors qu'une machine thermique, destinée à refroidir le sang surchauffé par les émotions ressenties par le cœur. [René Misslin, 2008] Pendant plus que mille ans les connaissances sur le cerveau n'évoluèrent guère, par un manque d'outils d'exploration efficace d'une part, et par doctrine religieuse qui interdisait toute étude anatomique humaine d'autre part, sous peine d'excommunication.

C'est Durant la Renaissance (seizième siècle) que débuta la révolution anatomique avec en particulier Vesale (1514-1564), ouvrant la voie à la dissection humaine et à l'anatomie descriptive. Il fut à l'origine de la première descriptive rigoureuse illustrée du corps humain, et du cerveau (Figure 1). Les découvertes de la cellule par Théodore Schwann (1840), Puis des cellules du système nerveux, les neurones par Golgi (1843-1926) et Cajal (1852-1934) grâce à une nouvelle technique de coloration furent fondamentales pour les études sur la forme, les propriétés, les fonctions et les connexions des neurones.

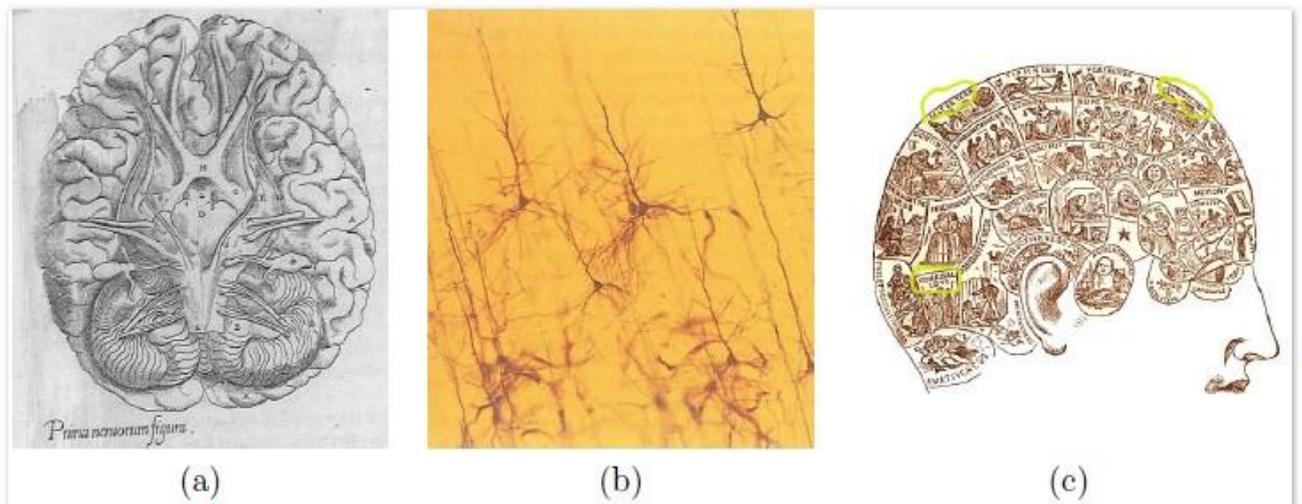


Figure 1.1 – Atlas de Vesale (image a) (source : De numani corporis fabrica), neurones colorés par la méthode de Golgi (image b) (source : Hubel) et phrénologie de F.J. Gall.

Depuis le début du vingtième siècle, toutes les grandes révolutions scientifiques et techniques, d'électronique à la mécanique quantique, de la génétique à l'informatique, ont trouvé application dans l'étude du cerveau.

Particulièrement, la découverte des rayons X en 1895 par Wilhelm Röntgen permet d'explorer le corps humain. Il n'était alors plus nécessaire d'avoir recours à la dissection pour explorer l'intérieur du corps, Au début des années 1970 les premiers ordinateurs assez puissants pour réaliser la reconstruction en 3D de l'image ont permis la mise au point du scanner (Hounsfield, 1970). L'utilisation d'ultrasons (échographie), puis l'imagerie par résonance magnétique (IRM). Ces méthodes d'Imagerie cérébrale permettent d'observer des aspects complémentaires de l'anatomie ou de l'activité cérébrale. Accompagné par l'évolution de l'informatique, elles ont entraîné une véritable révolution dans la connaissance du fonctionnement du cerveau à des fins d'aide au diagnostic des tumeurs cérébrales.

Les nombreux développements en physique et en informatique ont permis l'observation de plus en plus fine de différents aspects du cerveau. Les nouvelles techniques disponibles aujourd'hui essayent de mieux caractériser les différentes parties du cerveau. Parmi les familles de caractérisation, on trouve la texture des images qui permet d'avoir une source énorme d'information pour les utiliser ensuite dans ce qu'on appelle la classification des tumeurs cérébrales. L'objet de cette thèse est de proposer une nouvelle approche permettant d'orienter ou bien d'assister les experts (médecins radiologues) dans la reconnaissance des tumeurs cérébrales qui représentent aujourd'hui une des plus grandes causes de mortalité dans la société.

## Introduction générale

Aujourd'hui, la masse d'information produite par l'imagerie médicale dans un contexte globale, a considérablement augmentée. Cela est dû à l'apparition des nouvelles modalités d'acquisitions, qui permettent de visualiser tous ce qui est à l'intérieur du corps humain, nous pouvons citer par exemple le scanner à rayons X, Tomographie par émission de positons (TEP), l'échographie et l'imagerie par résonance magnétique (IRM). Cette dernière a connu un véritable essor ces dernières années, grâce à ses caractéristiques : une technique non invasive, permet en effet une observation de qualité de l'anatomie cérébrale. L'image obtenue, de haute résolution, offre une localisation précise des différentes structures. Cette modalité d'imagerie est ainsi devenue un outil de plus en plus central dans la médecine du cerveau et plus particulièrement dans le cadre du diagnostic des pathologies cérébrales.

Le médecin doit analyser des images médicales afin de diagnostiquer certaines maladies liées à des lésions cérébrales. Pour étudier l'évolution d'une tumeur, il est nécessaire de connaître avec exactitude les changements survenus sur ces images. Cependant Il n'est pas toujours évident de délivrer, même pour un œil expert, un diagnostic fiable à la simple vue de ces images.

Dans un souci d'aide aux médecins, le concept de systèmes experts informatiques dédiés à l'aide au diagnostic (CAD : Computed Assisted Diagnosis), émerge depuis une quinzaine d'années. Plusieurs équipes se sont intéressées aux développements d'outils, pour l'aide à la détection et au diagnostic des différentes lésions. Ils sont représentés par un outil informatique qui vise à aider mais non pas à remplacer le clinicien. Leurs principes se résument en trois étapes, à savoir la numérisation (effectuer dans la majorité des cas par la machine d'acquisition), l'extraction de caractéristiques propres aux images, la classification et le diagnostic.

C'est pour cela que le besoin d'une interprétation automatique ou semi-automatique qui permet d'assister les médecins dans leur prise de décision s'est fait ressentir. Autrement dit le traitement des images médicales est une étape cruciale dans le diagnostic des lésions cérébrales et plus précisément les tumeurs. L'une des propriétés des images médicale qui facilite cette tâche est la texture, une composante clé qui représente une riche source d'informations visuelles. Donc l'exploitation de la texture est nécessaire pour avoir des caractéristiques sur l'image médicale, plus précisément la tumeur.

La classification, représente l'étape finale dans le processus d'interprétation automatique ou semi-automatique, le but est de prédire les catégories de l'image d'entrée à l'aide des attributs qui caractérise l'image.

Dans ce mémoire, nous nous intéressons aux diverses méthodologies et techniques développées dans la littérature pour la caractérisation et la classification des images IRM. Afin de bien reconnaître la tumeur.

## **Problématique**

Le mot « Cancer » est utilisé pour désigner de nombreuses maladies différentes ayant pour point commun la présence de cellules cancéreuses. La plupart du temps, ces cellules cancéreuses forment des tumeurs dites « malignes », certaines, dites « Bénignes ».

Le cancer est devenu l'une des premières causes de décès dans le monde et spécifiquement celui touchant le cerveau. Donc la détection et l'élimination rapide et immédiate des tumeurs cérébrales est indispensable, afin de prolonger la durée de vie d'un patient.

Parmi les solutions qui existent dans le terrain est celle de l'imagerie médicale, et en particulier l'imagerie par résonance magnétique, qui permet d'observer l'activité de la tumeur illustrée dans des images. Les médecins trouvent parfois des difficultés dans l'interprétation de ces images pour donner un diagnostic fiable sur l'état du patient.

A cet effet, vu la difficulté à détecter la nature des lésions, nous nous intéressant dans ce travail à proposer aux médecins un système d'aide au diagnostic des tumeurs de cerveau en les classant en deux classes (Maligne et Bénigne) selon leurs textures.

## **Organisation du manuscrit**

Nous avons structuré le manuscrit en quatre chapitres ordonnés comme suit :

Le chapitre 1 introduit les termes et les concepts essentiels dans un contexte médical, nous présentons dans un premier temps quelques notions élémentaires de l'anatomie cérébrale (Section 1.2), nous donnons un rapide aperçu sur les tumeurs cérébrales (causes, symptômes et classification selon l'OMS) ensuite leurs diagnostics. Nous nous focalisons sur l'un des types d'imagerie médicale, Imagerie par Résonance Magnétique (IRM) et nous terminons par les nouveautés et l'actualité dans le domaine de traitement des tumeurs cérébrales.

Le chapitre 2 vise à établir un état de l'art sur les différentes méthodes d'extraction d'attribut de textures retenues dans la littérature, pour la caractérisation et la classification des images IRM cérébrales, avec les fondements mathématiques de ces méthodes.

Dans le chapitre 3 nous détaillons les méthodes de caractérisation et classification que nous avons utilisée. Nous commençons par introduire une méthode statistique d'ordre deux, à savoir des matrices de Cooccurrences (MC), nous présentant ensuite notre deuxième méthode de caractérisation qui est l'Analyse fractale. Enfin, nous terminons par abordant le cadre général de notre méthode de classification, Les Forêts Aléatoires, en anglais « Random Forest »

Le chapitre 4 décrit l'essentiel de notre travail, nous commençons par la phase de prétraitement des images IRM de la base de données BRATS, 2015, ensuite nous allons présenter la procédure de construction des bases de données utilisés dans la classification, et nous continuons par exposer les résultats obtenus avec l'interprétation de ces résultats. Nous terminons ce chapitre par un aperçu sur l'application développé en JAVA FX. Enfin, nous terminons ce modeste travail par une conclusion générale.

# CHAPITRE 1

## CONTEXTE MEDICALE

## 1.1 Introduction

Le cerveau humain, une masse de tissu spongieux d'environ 1300 g - est la structure vivante la plus complexe de l'univers. Capable de créer un réseau de connexions qui dépasse de loin n'importe quel réseau social, capable de stocker plus d'informations qu'un superordinateur, le cerveau permet aux êtres humains de penser, apprendre, et de créer. Pourtant, les scientifiques sont loin d'avoir découvert tout ce que le cerveau peut faire. Ce seul organe contrôle tous les aspects de notre corps, aussi bien les battements cardiaques que l'activité sexuelle, les émotions, l'apprentissage et la mémoire. Ce sont les capacités du cerveau à réaliser cet ensemble de fonctions qui font de l'être humain ce qu'il est.

Il n'est pas exagéré de souligner l'importance des recherches en neurosciences. Plus d'un millier de maladies du cerveau et du système nerveux sont responsables de plus de temps d'hospitalisation qu'aucun autre type de maladie. Donc l'exploration et la compréhension de cet organe est primordiale. [16]

Des techniques d'imagerie révolutionnaires, y compris l'imagerie par résonance magnétique (IRM), ont permis d'explorer le corps humain pour pouvoir examiner les différentes parties du cerveau afin d'assister les experts pour un diagnostic avancé.

Dans ce chapitre nous vous invitons à découvrir dans un premier temps l'anatomie du cerveau, nous présentons, ensuite les différentes lésions et tumeurs, ainsi leurs diagnostics, et nous terminons avec les techniques et nouveaux traitements utilisées de nos jours, en particulier l'imagerie par résonance magnétique qui donne une vue précise de l'intérieur du corps humain, avec une conclusion générale sur ce chapitre.

## 1.2 Un peu d'anatomie pour comprendre le cerveau

Situé à l'intérieur du crâne, le cerveau gère et coordonne la plupart de nos fonctions : les fonctions internes du corps (battements du cœur, circulation du sang, contractions des muscles ou digestion), les fonctions dites supérieures (pensée, émotions, personnalité, capacité à communiquer ou à apprendre) et enfin, les cinq sens, qui nous permettent d'être en relation avec l'extérieur : la vue, l'ouïe, le toucher, l'odorat et le goût.

Grâce aux progrès de la science, le fonctionnement du cerveau est de mieux en mieux connu. Nous savons aujourd'hui qu'il fonctionne grâce à des milliards de neurones, connectés entre eux. Nous savons également qu'il est organisé en plusieurs zones, qui gèrent chacune des fonctions différentes.

Cette connaissance du cerveau a permis de mieux comprendre les maladies dont il peut être atteint et de développer des traitements de plus en plus précis et efficaces.

### 1.2.1 Que se passe-t-il à l'intérieur du cerveau ?

Le cerveau fonctionne comme un ordinateur central. Il reçoit en permanence une multitude d'informations sur ce qui se passe à l'intérieur et à l'extérieur du corps. Ces informations lui proviennent des nerfs rachidiens, qui parcourent l'ensemble du corps, et des nerfs crâniens, situés à l'intérieur du crâne.

Chaque information est classée, analysée et traitée par des cellules nerveuses, les neurones. Ces derniers sont connectés entre eux et ils sont capables de gérer plusieurs informations à la fois. C'est ce qui nous permet de voir, de parler, de marcher et de respirer, tout cela en même temps et sans avoir à y réfléchir.

Ces milliards de neurones sont entourés et nourris par d'autres cellules appelées cellules gliales. Les cellules gliales sont dix fois plus nombreuses que les neurones. Elles représentent 90 % de la composition du cerveau. Il en existe différents types : les astrocytes, les oligodendrocytes, les épendymocytes et les cellules microgliales.

Un peu plus de la moitié des tumeurs du cerveau se développent à partir de ces cellules. Ce type de tumeur est appelé gliome. [3]

## 1.2.2 Les différentes parties du cerveau et leurs fonctions

Le cerveau est très organisé. Il est composé de plusieurs parties qui ont chacune des rôles spécifiques, tout en étant complémentaires les unes des autres. Parmi ces parties :

### Les hémisphères cérébraux

Les hémisphères cérébraux forment les parties les plus volumineuses du cerveau. Ils constituent la partie la plus importante du cerveau. Ils sont le siège de la raison et de la créativité. Ils contrôlent l'ensemble de nos fonctions mentales supérieures : mouvements volontaires, pensée, apprentissages, mémoire, etc.

D'une manière générale, l'hémisphère droit commande le côté gauche du corps et inversement.

Chaque hémisphère est lui-même partagé en quatre zones appelées lobes dans lesquelles différentes fonctions sont gérées (Figure 1.2)



Figure 1.2 Les quatre lobes de l'hémisphère gauche

### Le tronc cérébral

Le tronc cérébral relie les hémisphères cérébraux à la moelle épinière. C'est lui qui contrôle les fonctions vitales du corps : battements du cœur, respiration, tension artérielle. [4]

## Le cervelet

Le cervelet est situé à l'arrière du tronc cérébral, sous les lobes occipitaux. Il nous permet d'avoir des réflexes, de coordonner nos mouvements et de garder l'équilibre. [4]

### 1.2.3 Le cerveau, un organe sous haute protection

En tant que centre de commande de l'ensemble du corps, le cerveau bénéficie de trois systèmes de protections (Figure 1.4) :

- Les os du crâne forment une première coque solide autour du cerveau. Ils le protègent des chocs ;
- Trois membranes, les méninges, enveloppent le cerveau et la moelle épinière. Elles les protègent des blessures et des infections ;
- Enfin, à l'intérieur du crâne et le long de la colonne vertébrale, un liquide joue le rôle d'isolant et d'amortisseur contre les chocs. Ce liquide est le liquide céphalo-rachidien (abrégé en LCR), appelé aussi liquide cérébro-spinal (abrégé en LCS). Il est produit à l'intérieur du cerveau, dans des cavités appelées ventricules. [5]

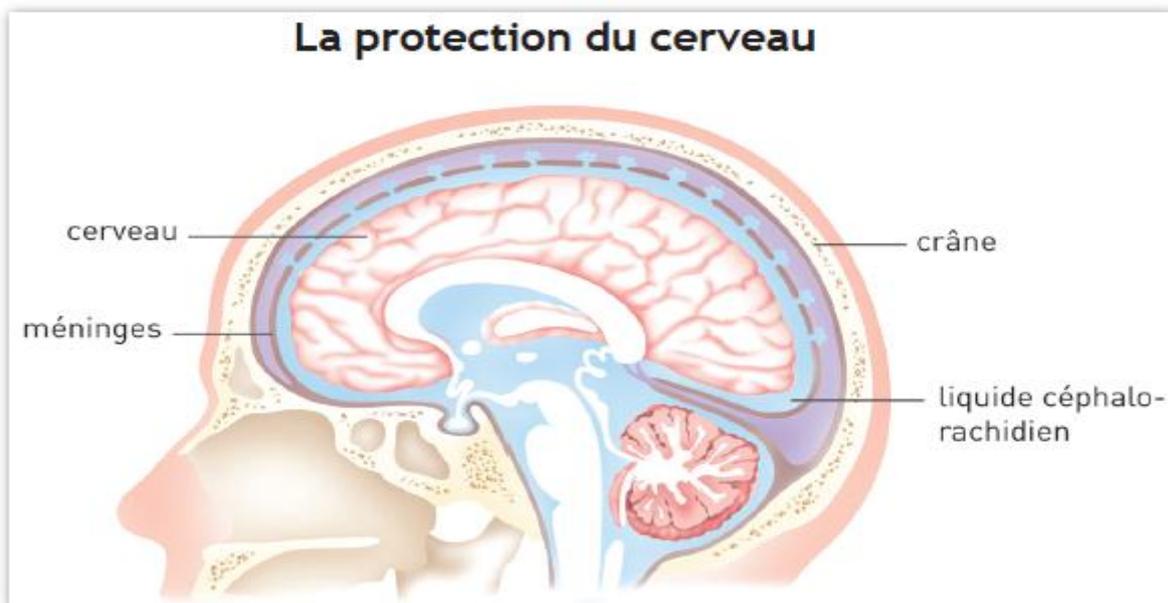


Figure 1.3 La protection du cerveau

Malgré cette protection, le cerveau se confronte d'une tumeur considérée comme la plus agressive qui affecte le corps humain.

## **1.3 Les tumeurs cérébrales**

### **1.3.1 Qu'est-ce qu'une tumeur ?**

Une tumeur est une masse plus ou moins volumineuse due à une multiplication anormale de cellules.

La cellule est l'unité de base dont sont constitués tous les tissus des organismes vivants. Il en existe dans le corps plus de deux cents types différents : les cellules musculaires, nerveuses, osseuses, etc.

Chaque cellule a un rôle précis et une durée de vie limitée. En permanence, les cellules vieillissent, meurent et sont remplacées par d'autres.

Et pour des raisons encore inconnues, il arrive que certaines cellules se modifient et continuent à se multiplier au lieu de disparaître naturellement. Elles se multiplient alors jusqu'à former une tumeur.

D'une manière générale, il existe trois types de Tumeurs [3] :

#### **Les tumeurs non cancéreuses, appelées tumeurs bénignes.**

Elles se développent lentement et restent localisées. Une fois traitées, elles ne récidivent pas généralement.

#### **Les tumeurs cancéreuses, appelées tumeurs malignes.**

Elles se développent plus rapidement et ont tendance à envahir d'autres zones que celles où elles sont apparues au départ. Elles peuvent développer alors de nouvelles tumeurs appelées Métastases.

#### **Les tumeurs intermédiaires dites « atypiques » ou « évolutives ».**

Bénignes au début, elles peuvent se transformer en cancer dans un laps de temps variable.

Avant de quitter cette partie, il est indispensable de marquer le point sur la différence entre les tumeurs primaires et secondaires. Les tumeurs du cerveau qui naissent directement à l'intérieur du crâne dite primaires ou primitives, Les métastases au cerveau

sont des tumeurs dites secondaires, qui se développent à partir d'un cancer situé dans une autre zone du corps humain : les poumons, le foie, le sein, etc.

### **1.3.2 Tumeur de cerveau**

On appelle tumeur du cerveau toutes les tumeurs qui se développent à l'intérieur du crâne. Elles peuvent se développer dans n'importe quelle zone du cerveau : les hémisphères, le cervelet, le tronc cérébral, l'hypophyse. (Voir « Un peu d'Anatomie pour comprendre le Cerveau », section 1.2)

Il existe des dizaines de tumeurs cérébrales différentes, que l'on distingue en fonction de trois caractéristiques :

#### **Selon la localisation dans le cerveau**

Comme nous l'avons indiqué précédemment, le cerveau est organisé en plusieurs zones, qui gèrent chacune des activités spécifiques : le langage, l'équilibre du corps, les battements du cœur, etc. Une tumeur peut entraîner des troubles qui se différencient selon la zone dans laquelle elle se développe. La localisation de la tumeur est également un élément essentiel pour le choix de traitements. [3]

#### **Selon le type**

Ici on parle de type histologique, c'est-à-dire les tumeurs du cerveau portent généralement le nom des cellules à partir desquelles elles se développent.

L'OMS distingue environ 200 types de tumeurs cérébrales, classifiées en fonction des cellules dont elles sont issues ; [4]

Les tumeurs cérébrales les plus fréquentes sont :

- Les gliomes (58%), principalement : astrocytomes, glioblastomes, oligodendrogliomes, épendymomes ;
- Les méningiomes (20%) ;
- Les adénomes de l'hypophyse (14%) ;
- Les neurinomes (7%) ;
- Les médulloblastomes ;
- Les lymphomes.

Les métastases ne sont pas prises en compte dans cette énumération. Elles représentent le 30-40% de toutes les tumeurs intracrâniennes.

### **Selon le degré d'agressivité**

Plus la tumeur se développe rapidement, plus elle est jugée agressive. Les tumeurs du cerveau sont classées en différents grades selon leur agressivité. On parle de bas grade pour les tumeurs les moins agressives et de haut grade pour les tumeurs qui le sont davantage.

Le grade peut aussi être exprimé par un chiffre romain allant de I à IV. Le grade I correspond aux tumeurs non cancéreuses, qui sont les moins agressives. Le grade IV correspond aux tumeurs les plus agressives (Voir tableau 1.1).

### **1.3.3 Les causes des tumeurs de cerveau**

Les causes des tumeurs du cerveau sont mal connues. On ne sait pas pourquoi, à un moment donné, les cellules se multiplient de manière incontrôlée jusqu'à former une tumeur.

De nombreuses études scientifiques ont été menées ou sont en cours, pour tenter de déterminer les facteurs qui favorisent l'apparition d'une tumeur au niveau du cerveau. On parle de facteurs de risques. Trois types de facteurs de risques sont étudiés : les risques liés l'environnement, à l'hérédité et aux virus. [3]

Un facteur de risque n'explique pas à lui seul pourquoi une tumeur est apparue. En effet, pour deux personnes confrontées au même risque, l'une peut développer une tumeur et l'autre non. À l'inverse, une personne peut être atteinte d'une tumeur alors qu'elle n'est concernée par aucun facteur de risque.

### **1.3.4 Les symptômes d'une tumeur de cerveau**

Les symptômes provoqués par une tumeur du cerveau sont très variables et n'apparaissent pas systématiquement. Ils dépendent du volume de la tumeur, de la vitesse à laquelle elle se développe et surtout de son emplacement. Certaines tumeurs ne provoquent aucun symptôme, ce qui signifie généralement qu'elles se développent lentement.

Trois types de symptômes sont possibles :

- Des maux de tête (céphalées), liés à une augmentation de la pression à l'intérieur du crâne ;
- Des crises d'épilepsie, liées à un dérèglement de l'activité des neurones ;
- Des troubles fonctionnels, directement liés à la localisation de la tumeur et aux fonctions gérées dans cette zone du cerveau. Ce peut être des troubles de la vision et de l'ouïe, des modifications de la personnalité, des difficultés à coordonner ses mouvements ou à trouver ses mots, des troubles digestifs (nausées, vomissements et perte d'appétit) et la faiblesse. [3]

### 1.3.5 Gradation des tumeurs selon l'OMS

Des grades établis par l'OMS (Organisation mondiale de la santé) permettent une classification des tumeurs en fonction de leur malignité et de leurs caractéristiques. [4]

Grade I Tumeurs Bénignes	Tumeurs à croissance lente et circonscrites (limites bien définies).
Grade II Tumeurs Bénignes	Tumeurs à croissance lente, mais, contrairement aux tumeurs de grade I, leurs limites sont imprécises.
Grade III Tumeurs Malignes	Tumeurs anaplasiques Un ou plusieurs foyers «anaplasiques» se sont développés au sein d'une tumeur préexistante de bas grade (I ou II). Leur croissance est plus rapide que celle des tumeurs de bas grade.
Grade IV Tumeurs Malignes	Les tumeurs de grade IV présentent une anaplasie très prononcée. Leur croissance est plus rapide que celle des tumeurs de bas grade.

Tableau 1.1 - Gradation des tumeurs selon l'OMS

Maintenant Lorsqu'une tumeur du cerveau est découverte, il est indispensable de déterminer ses caractéristiques : sa localisation, son type et son degré d'agressivité. Pour cela, plusieurs examens doivent être réalisés. Ce n'est qu'après ces différents examens que l'on sait de quel type de tumeur il s'agit et que l'on peut définir les traitements appropriés.

## **1.4 Diagnostic d'une tumeur de cerveau**

Généralement, c'est l'apparition des symptômes qui permet de suspecter la présence d'une tumeur du cerveau.

Plusieurs étapes sont nécessaires pour déterminer l'origine des symptômes décrits et confirmer ou infirmer la présence d'une tumeur cérébrale.

### **1.4.1 Examen clinique et neurologique**

L'examen clinique et neurologique est réalisé au cours d'une consultation médicale, soit par le médecin traitant, soit par un médecin spécialiste du cerveau. Il consiste à évaluer l'état de santé général d'une part et le bon fonctionnement du système nerveux d'autre part.

L'examen clinique et neurologique permet de constater d'éventuelles anomalies. Pour déterminer si ces anomalies sont dues à une tumeur ou non, des examens d'imagerie sont indispensables.

### **1.4.2 Les examens d'imagerie**

Les examens d'imagerie consistent à réaliser des images précises du cerveau, afin de confirmer ou infirmer la maladie. En cas de résultat positif, ils permettent de détecter une tumeur, de la localiser de façon précise, de mesurer sa taille et d'évaluer les conséquences qu'elle a ou peut avoir sur le cerveau.

Pour le diagnostic d'une tumeur du cerveau, une IRM est indispensable. Pour des raisons pratiques, l'IRM est souvent précédée d'un scanner. [3]

Dans notre travail nous nous intéressons aux images cérébrales acquises par l'imagerie par résonance magnétique connu sous le nom d'IRM qu'on va le présenter par la suite. (Section 1.5)

### **1.4.3 Analyser un échantillon de tumeur : L'examen anatomopathologique**

Après les examens d'imagerie, l'analyse d'un échantillon de la tumeur est indispensable pour compléter le diagnostic. Cette analyse est appelée examen anatomopathologique.

L'objectif de l'examen anatomopathologique est de déterminer toutes les caractéristiques de la tumeur : sa nature exacte, sa composition, son degré d'agressivité. C'est seulement après cet examen que les médecins peuvent affirmer notamment si la tumeur est cancéreuse ou non et quel est son grade.

Dans des cas très rares, notamment chez des personnes âgées, l'examen anatomopathologique ne peut pas être réalisé, car aucun prélèvement n'est possible. Les médecins peuvent alors, exceptionnellement, décider de traiter directement la tumeur par une radiothérapie ou une chimiothérapie.

## **1.5 Imagerie par résonance magnétique**

Depuis la toute première radiographie, en 1895, l'imagerie médicale a fait des progrès saisissants. Révolutionnant la médecine, elle permet aujourd'hui de visualiser les os, les tissus et les organes de façon toujours plus précise et moins invasive. Le but ? Dépister les anomalies, évidemment, mais aussi, et de plus en plus, soigner.

Cette technique magique d'imagerie médicale basée sur le phénomène de résonance magnétique nucléaire (RMN, terme créé en 1930 par Isidor Isaac Rabi ; prix Nobel de Physique en 1944) a connu un véritable essor ces dernières années. L'IRM anatomique est aujourd'hui la méthode non invasive la plus adaptée pour observer avec une résolution fine les tissus biologiques, et donc en particulier les tissus cérébraux.

### **1.5.1 Principe de fonctionnement**

Une IRM est réalisée grâce à un grand appareil en forme de cylindre. Cet appareil est composé d'un aimant très puissant (d'où le terme de magnétique). Il produit des ondes radio qui sont projetées sur le cerveau et permettent d'obtenir des images « en coupe ».

Les images sont ensuite assemblées par un ordinateur pour obtenir une reproduction très précise du cerveau.

Pendant l'examen, un produit de contraste est injecté dans une veine du bras. Il permet de mettre en évidence certains aspects du cerveau, comme les vaisseaux sanguins et facilite l'interprétation des images.

### **1.5.2 Avantages et inconvénients**

Avantages :

- Précision, innocuité, prises de vue dans toutes les orientations, 3D intrinsèque ;
- Technique non irradiante ;
- Très forte résolution spatiale (plus de détails par rapport au scanner).

Inconvénient :

- Relative lenteur (une image en quelques minutes) ;
- Un examen bruyant, les patients met des bouchons d'oreille pour réduire le bruit ;
- Coût élevé.

## **1.6 Technologie et nouveaux traitements**

La recherche médicale et scientifique a permis récemment de développer de nouveaux traitements appelés traitements ciblés ou thérapies ciblées. Le principe de ces traitements est de priver la tumeur des éléments dont elle a besoin pour se développer : oxygène, nutriments, vaisseaux sanguins...

En ciblant mieux la tumeur, ces médicaments pourraient améliorer l'efficacité des traitements des tumeurs du cerveau, et diminuer leur toxicité.

Les traitements énumérés ci-dessous sont encore au stade expérimental et rien ne permet à ce jour de dire quelle place ils occuperont à l'avenir.

### **Le traitement génétique**

Les recherches dans le domaine de la thérapie génétique se concentrent principalement sur deux axes. D'une part l'introduction, dans la tumeur, d'un gène qui empêcherait la multiplication des cellules cancéreuses et d'autre part, l'introduction d'un gène qui rendrait les cellules sensibles à un traitement médicamenteux.

### **La vaccination**

Des recherches sont également en cours pour la mise au point d'un vaccin à base de cellules tumorales.

## Les thérapies antiangiogéniques

La néoangiogénèse, c'est-à-dire le développement de nouveaux vaisseaux sanguins, joue un rôle fondamental dans la croissance tumorale. Les vaisseaux sanguins sont indispensables à la croissance de la tumeur car ils l'alimentent en oxygène et en nutriments.

Les thérapies antiangiogéniques ont pour objectif d'empêcher la formation de nouveaux vaisseaux sanguins afin «d'asphyxier» les tumeurs.

Maintenant nous abordons rapidement dans la suite une nouvelle technique d'Imagerie cérébrale qui vient de remplacer l'IRM, comment ? Et Pourquoi ?

### **Le challenge : une technique d'imagerie cérébrale portable, robuste et ultraperformante : Le FUltraSound**

Mickael Tanter, directeur de recherche à l'Inserm et ses collaborateurs (équipe « Physique des ondes pour la médecine » de l'Institut Langevin, CNRS/ESPCI Paris Tech) publient dans la revue « *Nature Methods* », un article présentant une nouvelle technique d'imagerie très prometteuse : le FultraSound.

Une nouvelle technique doit donc être inventée pour pouvoir réaliser des images en 3 dimensions du cerveau d'animaux éveillés et libres de leurs mouvements, qui aurait le potentiel d'être transposée par la suite à l'homme, et qui présenterait des performances supérieures à celles de l'IRM pour être adoptée par les médecins cliniciens.

Le principe est la mise au point d'une nouvelle technique d'imagerie cérébrale ayant justement ce potentiel : elle met en œuvre des sondes ultrasonores légères (et donc portables) et elle opère à une fréquence assez élevée pour s'affranchir des mouvements cardiaques et respiratoires (qui se signalent à des fréquences plus basses). Elle permet de réaliser des images de la connectique fonctionnelle du cerveau d'animaux vivants tels que le rat, avec une précision spatiale, une résolution temporelle et une sensibilité de détection bien supérieures à celles de l'IRM. [17]

## 1.7 Conclusion

Au terme de ce chapitre, nous avons survolé sur les diverses modalités médicales qui ont une forte relation avec notre thème, qui nécessite en premier temps une description anatomique du chef d'orchestre dans l'organisme humain « Le Cerveau », puis nous avons décrit la maladie la plus dangereuse du cerveau, qui a malheureusement des effets indésirables sur d'autres organes et l'ensemble d'organisme, on parle de la tumeur cérébrale. Ensuite nous avons essayé d'aborder les différentes techniques utilisées dans la phase de diagnostic, en particulier l'imagerie médicale qui a révolutionné les connaissances sur le cerveau notamment l'imagerie par résonance magnétique (IRM) qui reste le premier choix dans l'exploration du corps humain et principalement l'imagerie cérébrale.

Grace à ces techniques d'imagerie médicale, la masse d'information produite a considérablement augmentée, donc l'interprétation automatique des IRM cérébrales est devenue un enjeu majeur. Et les médecins ont besoin d'outils fiables pour les assister dans leur prise de décision et dans l'interprétation de la masse d'information créée...

Alors comment faire pour les aider ? Nous vous invitons d'aller au chapitre suivant où on va présenter les différentes techniques de caractérisation d'images IRM du cerveau, afin d'extraire l'information nécessaire pour la classification des tumeurs cérébrales.

# CHAPITRE 2

## ETAT DE L'ART SUR LES METHODES D'EXTRACTION D'ATTRIBUT DE TEXTURE

## 2.1 Introduction

L'analyse de textures joue un rôle très important. Elle permet l'identification des objets. En effet, nous verrons qu'il est inévitable, pour distinguer les différentes régions d'une image de même couleur (ou bien de même niveau de gris), d'extraire une information texturale. Malgré cette importance, nous verrons tout d'abord qu'il est difficile de caractériser précisément la texture.

Nous essayerons néanmoins d'en donner une définition générale. Ensuite, dans un deuxième paragraphe, nous présenterons les différentes méthodes d'analyse de textures. Nous mettrons en évidence les principales méthodes d'extraction de paramètres permettant d'identifier les textures. L'accent sera mis sur l'utilisation des matrices de cooccurrence et les méthodes fractales, car elles offrent une bonne caractérisation de la notion de texture.

Comme le but de notre analyse est d'extraire et de tester des paramètres caractérisant les textures, nous présenterons les paramètres les plus utilisés dans différentes techniques. Enfin nous terminons avec une introduction sur la classification.

## 2.2 La notion de texture

### 2.2.1 Définition

Revenons sur la notion de texture. Celle-ci est en fait difficile à définir quantitativement ; elle caractérise l'organisation et la structure d'une région macroscopique d'une image. Nous retiendrons la définition suivante, proposée par [Haralick, 1979] :

**Définition** : La texture peut être décrite comme une structure hiérarchique à deux niveaux. Le premier concerne la description d'éléments de base ou primitives, à partir desquels est construite la texture. Le second niveau est relatif à la description de l'organisation spatiale de ces primitives, à l'agencement des motifs texturaux entre eux, qui peut satisfaire aussi bien à des lois stochastiques que déterministes.

En imagerie biomédicale, il n'existe pas de véritable consensus en ce qui concerne la définition d'une texture. Dire *qu'une texture est une région d'une image présentant une organisation spatiale homogène des niveaux de luminance*.

### 2.2.2 Type de textures

On distingue en fait, selon l'élément de base et les règles de placement, 2 types de textures :

#### Les textures structurées

Les **Macrotextures** (ou textures structurées) pour lesquelles il est facile d'extraire visuellement le motif de base et les lois d'assemblage des primitives entre elles. Ces textures peuvent même présenter une certaine périodicité ou cyclostationnarité. [35]

#### Les textures aléatoires

Les **Microtextures** (ou textures aléatoires) qui présentent un aspect plus chaotique et plus désorganisé, mais dont l'impression visuelle reste globalement homogène. [35]

## 2.3 Etat de l'art des différentes méthodes d'extraction d'attribut de texture

Un pixel est caractérisé par son emplacement sur l'image et son intensité (couleur, niveau de gris...). Qu'en est-il donc d'une texture ?

Tout d'abord il faut distinguer deux types de textures : Les macro-textures et les micro-textures. Cet état de l'art se consacre à l'étude de ce second type de textures. Dans ce cas, un élément de texture peut être considéré comme un ensemble de pixels voisins les uns des autres, caractérisés par la manière dont sont distribuées les intensités de ces pixels.

Donc, Il n'est pas possible de localiser précisément une texture car ses contours sont flous. De même elle est intuitivement décrite par sa finesse (ou grossièreté), sa régularité, sa douceur ou bien même son orientation. Pour l'étudier de manière formelle, il est donc nécessaire de retraduire ces descripteurs sous forme de formules. [14]

De nombreuses méthodes permettent d'extraire les caractéristiques des textures dans une image. Nous allons présenter dans cette partie l'état de l'art sur les méthodes les plus communément utilisées en analyse de la texture.

### 2.3.1 Méthodes statistiques

Les méthodes statistiques sont adaptées à l'étude des structures fines sans régularité apparente. Trois grandes classes de méthodes seront étudiées, celles du premier ordre, celles d'ordre deux et celles d'ordre supérieur. L'ordre d'une méthode étant ici le nombre de pixels mis en jeu pour l'évaluation de chaque résultat.

#### 2.3.1.1 Méthodes d'ordre un

L'analyse des régions d'une image par une méthode d'ordre un s'effectue directement à partir des niveaux de gris des pixels de l'image. Chaque pixel de la région sera donc représenté par une valeur commune à la région. Les paramètres d'étude les plus souvent utilisés sont :

$$\text{La moyenne : } MOY = \frac{1}{N} \sum_{i,j} g(i,j) \dots\dots\dots (1.1)$$

$$\text{La variance : } VAR = \frac{1}{N} \sum_{i,j} (g(i,j) - MOY)^2 \dots\dots\dots (1.2)$$

$$\text{L'aplatissement : } SKEW = \frac{1}{N} \sum_{i,j} (g(i,j) - MOY)^3 \dots\dots\dots (1.3)$$

$$\text{La dissymétrie : } KURT = \frac{1}{N} \sum_{i,j} (g(i,j) - MOY)^4 \dots\dots\dots (1.4)$$

Le rapport signal-bruit.  $SNR = \frac{MOY}{\sqrt{VAR}}$  ..... (1.5)

**Remarque**

Les méthodes du premier ordre permettent une étude simple et rapide mais ne procurent aucune information sur l'organisation locale des pixels (elles ne permettent pas d'étudier les relations entre deux Pixels). Des méthodes d'ordre supérieur sont donc utilisées afin de faire une analyse plus précise.

**2.3.1.2 Méthodes d'ordre deux**

**Matrices de Cooccurrences (MC)**

Cette méthode constitue la référence en analyse de texture et sert souvent comme étant une méthode comparative pour les nouvelles approches. Elle est, en effet, simple à mettre en œuvre et offre de bonnes performances.

En effet, chaque résultat met en valeur les relations entre deux pixels. Les matrices de Cooccurrences représentent donc les dépendances spatiales des niveaux de gris.

Soit une image ID (plus simplement notée I par la suite) à valeurs dans G : D est l'ensemble d'indices définissant le support de l'image I, ou d'une région de l'image, et G est l'ensemble des niveaux de gris accessibles. On note :

$$Ct = \{(\vec{x}, \vec{y}) \in D^2 | \vec{y} = \vec{x} + t \}$$

L'ensemble des couples de pixels séparés l'un de l'autre par une translation  $t \in Z^2$

**Remarque :** Les statistiques de cooccurrences sont les statistiques suffisantes d'un champ de Markov défini sur des cliques d'ordre 2

Si vous voulez en savoir plus sur les Matrices de Cooccurrences, et comprendre pourquoi cette méthode est devenue si importante dans l'extraction d'attributs de textures, abordez le chapitre suivant.

**2.3.1.3 Méthodes d'ordre supérieur**

**Matrices de longueurs de plage**

Les longueurs de plage (ou isosegments) est une méthode d'ordre supérieur. On appelle une plage un ensemble de pixels consécutifs et dans une direction donnée ayant le même niveau de gris. La longueur d'une plage est alors le nombre de pixels pour une plage donnée.

## Remarque

Il est utile de préciser que si certains paramètres correspondent à des propriétés visuelles de la texture, certains autres ne reflètent que des propriétés mathématiques auxquelles il n'y a pas forcément de propriété visuelle associée.

### 2.3.2 Méthodes structurelles

On considère que les textures sont constituées de motifs de base qui se répètent dans des directions de l'espace de façon quasi-régulière. Dans les images étudiées, on considère donc l'existence de primitives. Une description structurale d'une texture implique la recherche des motifs élémentaires, leur description, puis, la détermination des règles conditionnant leur position.

Plusieurs approches peuvent être regroupées en un ensemble de méthodes heuristiques, dites **méthodes structurelles** classiques, pour lesquelles on cherche à détecter et caractériser des primitives, puis, à trouver des règles de placement, (méthodes «bottomup») ou inversement (méthodes «top-down»). Elles s'appuient principalement sur le traitement du signal, la topologie et la géométrie. [19]

### 2.3.3 Méthodes spatio-fréquentielles

Représentations spatio-fréquentielles préservent à la fois les informations globales et locales ! bien adaptées aux signaux quasi périodiques. Permettent de caractériser la texture à différentes échelles. Et on trouve parmi ces méthodes :

#### Méthodes basées sur la Transformée de Fourier

La Transformée de Fourier permet, dans le domaine de l'image, d'obtenir une représentation fréquentielle spatiale des données. C'est-à-dire un passage de la représentation spatiale (espace géométrique 2D de l'Image) à la représentation fréquentielle.

## Méthodes de Bancs de filtres

Les méthodes basées sur des bancs de Filtres fournissent une autre approche pour l'analyse de texture, elles sont utilisées à la fois dans la classification et la segmentation. Filtres de Gabor [Bovik et al., 1990] est un exemple populaire d'une méthode basée sur le banc de filtres pour son invariance par rapport à l'échelle, la rotation et le déplacement.

## Transformée en Ondelettes

La décomposition en ondelette « DWT : DiscreteWavelet Transform » d'une image permet une analyse multi-résolution. Ce modèle procède à une analyse multi-échelles : une scène est d'abord perçue globalement puis dans les détails.

Un point de plus assez important pour la décomposition en ondelettes dans le domaine d'imagerie médicale et en particulier les images IRM d'une tumeur cérébrale ou [Oliviet, Sylvian, Valerie] : Ils sont arrivés à déterminer la différence entre les différents types de tumeurs (tumeurs High grade, bas grade,) se fait surtout grâce aux paramètres d'échelles (obtenue par la caractérisation du comportement des coefficients d'ondelettes sur  $2 \times 3$  sous-espaces d'ondelettes).

La multiplicité et la diversité des travaux sur la caractérisation des textures rendent difficile une catégorisation finie et exhaustive des différentes méthodes ; aussi, la suite donnera-telle un aperçu des approches « diverses » de la littérature.

### 2.3.4 Méthodes basées sur un modèle

#### Modèle de Markov

Utilisés depuis de nombreuses années, les champs aléatoires de Markov sont appliqués au traitement d'images. Leur utilisation, en analyse d'images, repose, essentiellement, sur le théorème d'Hammersley-Clifford qui donne une équivalence entre champ de Markov et distribution de Gibbs.

Considérons  $X_s$  la valeur du descripteur prise au site  $s$  et  $X^s = (X_t)_{t \neq s}$  la configuration de l'image excepté le site  $s$ . La définition d'un champ de Markov est alors la suivante :

$X$  est un champ de Markov ssi la probabilité conditionnelle locale en un site  $n$  est fonction que de la configuration du voisinage du site considéré

Ce qui s'exprime de façon formelle par :

$$P(X_i = x_i / x^l) = P(X_i = x_i / x_j; j \in V_i)$$

### Remarque

Les champs de Markov ne sont pas un outil d'étude de la texture à proprement parler. Ils sont plutôt un outil permettant de segmenter une image. Cependant, beaucoup de méthodes de segmentation d'images texturées sont fondées sur les champs de Markov.

### Modèle Fractale

Le mot fractal est un terme proposé par Mandelbrot (1975) qui vient du latin "fractus" et qui signifie irrégulier. Il caractérise toutes les formes géométriques qu'on peut imaginer comme faites de parties dont chacune est un modèle réduit du tout.

Généralement, une fractale est un objet composé de plusieurs sous-objets. La caractéristique globale de l'objet est similaire à la caractéristique locale de chacun des sous-objets.

La notion fractale est faite pour désigner des objets qui sont géométriquement compliquées (forme irrégulière, très interrompue) comme l'indique [8]

Tant qu'on parle sur la notion de Fractale il est indispensable de dire que la dimension fractale (DF) (mandelbrot 1984), qui est en fait utilisée dans divers domaines tel que la biologie, l'analyse d'écriture et l'analyse de texture, est un paramètre permettant de quantifier la complexité d'une image, caractérise le comportement auto-similaire de la surface, c'est un nombre qui quantifie le degré d'irrégularité et de fragmentation d'un ensemble géométrique.

Le problème ici est que les calculs s'effectuent sur des fenêtres de taille importante et demandent un grand temps de calculs.

L'avantage de cette approche réside dans le fait que la dimension fractale est assez peu sensible à la résolution de la texture et est fortement corrélée avec la perception humaine de la rugosité d'une surface ; elle permet de mesurer la complexité géométrique d'un ensemble.

Selon [5] la DF a montré des résultats intéressants, lorsqu'elle était appliquée sur des diverses modalités, telle que l'IRM, le scanner, et les ultrasons. Cependant, son efficacité prend tout son essor quant à elle est combinée avec d'autres attributs de texture.

Pour plus de détail sur la méthode d'Analyse fractale, nous vous invitons à poursuivre le chapitre prochain.

## **2.4 La classification**

La classification des données médicales est un domaine majeur de la Reconnaissance de Forme, elle fait appel à des modèles nommés « Classifieur ». Ces derniers ont besoin d'une phase d'apprentissage pour devenir autonome et intelligent, et après cette phase le classifieur va être testé sur des nouvelles données. La partie classification dépend énormément de la caractérisation des données. [11]

### **2.4.1 Définition**

La classification est une méthodologie consistant à classer des individus (des pixels, des images, . . .) en exploitant leur similarité, mais en aucun cas on ne cherche à partitionner l'image. Dans la classification d'images, les classes sont construites à partir d'attributs extraits des images, ou parties d'images, à analyser. Chaque image est assignée à une classe unique et est classifiée indépendamment des autres. [12]

### **2.4.2 Méthodes de classification**

Les méthodes de classification peuvent être divisées en deux catégories selon l'information disponible a priori sur les observations (Duda et Hart, 1973) (Jain, 1989) :

#### **Classification par apprentissage Supervisé**

Lorsque les différentes classes des observations sont connues au préalable, la classification est dite supervisée. Cette catégorie de méthodes est préférable pour la classification des images. Il existe une multitude de méthodes de classification supervisée ou les plus populaires sont : la règle des K plus proches voisins (anglais : KNN), Machine à vecteurs de support (SVM), Réseau de neurones (RN's) et Arbre de décision.

#### **Classification par apprentissage Non Supervisé**

En revanche, dans le cas où l'on n'a aucune connaissance a priori sur l'appartenance des observations aux classes, on parle de classification non-supervisée ou « *Clustering* », celle-ci est plus adaptée pour la segmentation des images. On cite parmi les méthodes de cette catégorie les plus connues comme : Classification Ascendante Hiérarchique, K-means.

Soulignons l'importance des méthodes de classifications à partir d'attributs de description de textures (Haralick, 1979).

La classification de textures consiste à associer une classe ou un label à une texture donnée, en fonction des caractéristiques présentes. Une première étape importante est donc la caractérisation ou bien l'extraction des caractéristiques de textures, ces attributs sont calculés à partir des méthodes qu'on va les aborder dans le chapitre suivant. Ces caractéristiques sont regroupées sous forme de vecteurs numériques et sont, par la suite, regroupées dans des bases de données, ces derniers seront directement utilisés par un Classifieur. Une deuxième étape importante de processus de classification est donc le choix du classifieur.

## 2.5 Conclusion

Dans ce chapitre, nous avons présenté les méthodes d'analyse des textures les plus usuelles avec un survole sur la notion de la classification, nous avons constaté que le choix de la méthode joue un rôle très important dans le processus et surtout pour avoir de bons résultats avec un temps de calcul rapide, c'est l'idéal.

Nous avons vu que parmi ces méthodes, les matrices de cooccurrence ou bien les longueurs de plage peuvent donner des bons résultats. Elles sont cependant assez lourdes en termes de calculs. Les méthodes d'ordre un sont beaucoup plus rapides. Cependant elles ne décrivent pas les relations spatiales entre les pixels.

Les méthodes fondées sur un modèle tel que le modèle Fractale ou les champs de Markov sont aussi largement utilisés en analyse de texture. Elles permettent d'obtenir de très bons résultats mais elles sont particulièrement lourdes en termes de calculs.

Dans le chapitre suivant, nous allons mettre le point sur nos méthodes d'extraction d'attributs de textures, on parle sur les Matrices de Cooccurrences et le Modèle Fractale avec un passage assez important sur notre méthode de classification « Les Forêts Aléatoires ».

# CHAPITRE 3

## NOS METHODES DE CARACTERISATION ET CLASSIFICATION

## 3.1 Introduction

L'imagerie est devenue un élément essentiel dans les diagnostics médicaux. Cette évolution est liée à la multiplicité des modalités d'acquisition ainsi qu'à l'exploitation informatique, qui en est faite.

L'analyse texturale fait partie de ces exploitations informatiques. La notion de texture ne peut être précisément définie [5] ; elle sert à qualifier mathématiquement des termes de langage utilisés pour quantifier les images, tels que : régularité, homogénéité, contraste... Dans le cadre des images médicales, ce genre d'analyse peut être utilisée pour diverses raisons :

- Distinction entre des zones saines et pathologiques.
- Aide au diagnostic sur l'évolution d'une pathologie.
- Caractérisation de zones informatives en vue d'opération de fusion...

Dans cette optique, plusieurs études ont été menées pour développer des algorithmes qui pouvaient quantifier les propriétés texturales d'une image. Cependant, durant ces dernières années, l'évolution incessante à la fois des algorithmes d'analyse de texture et des technologies ont ravivé l'intérêt des chercheurs pour des applications en imagerie médicale [lopes]. Lors de cette dernière décennie, les résultats des articles publiés ont montré la capacité de ces procédés à extraire des informations utiles pour le diagnostic d'images médicales.

Notre objectif est d'automatiser la reconnaissance de la tumeur en utilisant Une analyse quantitative visant à caractériser des zones tumorales par une analyse texturale, pour cela on va mettre le point sur deux méthodes d'analyse de texture, on parle, sur L'analyse Fractale et Les matrices de Cooccurrences d'où la nécessité d'introduire l'information texturale, pour ensuite réaliser une classification des images IRM (BRATS,2015) avec une technique qui s'appuie sur les arbres de décision de type CART.

Ce chapitre se divise en 2 grandes parties, la première partie traite nos différentes méthodes d'analyse de texture des images IRM, en se focalisant sur l'Analyse Fractale avec Les Matrices de Cooccurrences, pour extraire de l'information nécessaire pour la classification des tumeurs. La deuxième partie présente la classification avec « Les forets aléatoires »

## **Extraction d'attributs de textures**

Dans un souci d'aide aux médecins, le terme « d'aide au diagnostic » est très optimiste et reste encore aujourd'hui utopique. En effet, les techniques, mises au point, en sont plus à des procédés « d'aide à la détection », ceux qui constituent, à ce stade, des outils déjà très intéressants pour le clinicien, puisqu'ils permettent de cibler une zone dite « suspecte », que l'expert peut diagnostiquer.

Un des attributs utilisés dans ce procédé est la texture, laquelle est une riche source d'informations visuelles et est une composante clé dans l'analyse des images et la compréhension humaine.

Une signature est une représentation formelle des caractéristiques de la texture étudiée. Cette représentation est en générale caractérisée par une valeur numérique ou un vecteur de valeurs, on utilise souvent les termes attribut ou paramètres pour désigner l'ensemble de caractéristiques. [6]

Plaçons-nous maintenant dans le contexte de travail qui nous intéresse : étant donnée une base de données d'images textures, nous voulons les classifier selon un certain nombre de classes prédéfinies. Nous allons donc extraire pour chaque texture de notre base de données un vecteur caractéristique formé de paramètres caractéristiques de cette texture, et nous servir de ce vecteur pour la caractériser, c'est-à-dire l'associer à l'une ou à l'autre des classes.

Pour cela, nous vous invitons à découvrir pourquoi on a décidé de travailler avec les méthodes Fractales, Les Matrices de Cooccurrences pour la caractérisation et les Forêts aléatoires pour la classification des images IRM de la base BRATS, 2015

## **3.2 Méthodes d'analyse de texture, application aux images IRM cérébrales**

### **3.2.1 Méthodes basées sur les Matrices de Cooccurrences**

Du fait de leur richesse en information de texture, les approches basées sur les matrices de cooccurrence sont parmi les approches qui ont rencontré le plus grand succès en analyse de la texture. Appliquées initialement à la classification d'images, [Haralick, 1973]. Elle est, en effet, simple à mettre en œuvre et offre de bonnes performances.

Les Matrices de Cooccurrences font parties des méthodes statistiques du second ordre au même titre que les opérateurs de gradients. Leurs fonctions étant de mesurer les cooccurrences spatiales des niveaux de gris dans une région locale. Les paramètres issus des matrices de cooccurrences sont les plus utilisés en analyse de texture, notamment en analyse d'images médicales, où l'efficacité a été prouvée sur diverses modalités. Il existe plusieurs définitions de ces matrices, nous nous sommes basés sur celle proposée par Haralick. Nous définissons rapidement la méthode utilisée, des explications plus détaillées se trouvent dans la littérature. [Takagi and Shimoda, 1991]

Les approches statistiques, qui s'appuient sur la distribution des niveaux de gris des pixels négligent à priori l'aspect structurel des textures. Pour remédier à ce défaut, les matrices de cooccurrence permettent de décrire les textures à partir d'histogrammes bi variés directionnels, i.e. de la distribution conjointe des niveaux de gris des couples de pixels situés à des positions relatives données par un déplacement fixe (approche GLCM : Grey Level Cooccurrence Matrix) [Haralick et al. 1973]

### 3.2.1.1 Construction des matrices de cooccurrence

Nous considérons une image  $I$  avec  $G$  niveaux de gris. La matrice de Cooccurrences  $M$  sera donc de taille  $G * G$ , où chaque terme est une probabilité conjointe d'un couple de niveaux de gris séparés par une relation spatiale  $R$ .  $R$  est défini par un vecteur de déplacement  $r = (d, \theta, \varphi)$ , où  $d$  est la distance entre 2 voxels et  $(\theta, \varphi)$  la direction. Pour construire  $M$ , on va compter en  $M(i, j)$  le nombre de fois où le niveau de gris  $j$  à succéder au niveau de gris  $i$  pour un vecteur de déplacement  $r = (d, \theta, \varphi)$  fixé :

$$P_{d,\theta,\varphi} = |\{(I(u), I(u')) : I(u) = i, I(u') = j\}|$$

Où  $u = (x, y, z)$  et  $u' = (x', y', z') = (x + d \cos\theta \cos\varphi, y + d \sin\theta \cos\varphi, z + d \sin\varphi)$

Haralick et al. (Haralick et al. 1973) ont proposé de calculer à partir de ces matrices de cooccurrences des paramètres de texture pour quantifier la dépendance spatiale des valeurs de niveaux de gris. Il existe 14 paramètres. Chaque paramètre est censé apporter une information différente, par exemple le contraste mesure les variations locales des niveaux de gris. Si elles sont importantes (c'est-à-dire s'il existe peu de région homogènes), alors le contraste sera élevé. Cependant pour certaine texture, certains paramètres auront une information redondante par rapport à d'autres, cela peut se vérifier en calculant le coefficient de corrélation entre les paramètres. Il ne sera donc peut-être pas utile de les prendre tous en compte lors de nos applications.

### 3.2.1.2 Les indices de Haralick

Dans son article « Textural features for image classification », Haralick introduit quatorze attributs de texture extraits des matrices de cooccurrences. Ces attributs sont les suivants [12] :

1/ Second Moment Angulaire (Energie) : mesure l'homogénéité de l'image.

$$f_1^{k,k'} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{m^{k,k'} [I](i,j)\}^2$$

2/ Contraste (Inertie) : mesure les variations locales des couleurs.

$$f_2^{k,k'} = \sum_{n=0}^{N-1} n^2 \left\{ \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m^{k,k'} [I](i,j)}{|i-j|=n} \right\}$$

3/ Corrélation : Ce paramètre permet de déterminer si certaines colonnes de la matrice sont égales, c'est-à-dire s'il existe des dépendances linéaires dans l'image.

$$f_3^{k,k'} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu_x)(j-\mu_y) m^{k,k'} [I](i,j)}{\sigma_x \sigma_y}$$

4/ Variance : mesure la répartition des couleurs autour de la valeur moyenne.

$$f_4^{k,k'} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (m^{k,k'} [I](i,j) - \mu)^2$$

5/ Moment différentiel inverse : Ce paramètre a un comportement inverse de celui du contraste.

$$f_5^{k,k'} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2} m^{k,k'} [I](i,j)$$

6/ Entropie : mesure la complexité de l'image.

$$f_6^{k,k'} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m^{k,k'} [I](i,j) \log \{m^{k,k'} [I](i,j)\}$$

7/ Moyenne des sommes :

$$f_7^{k,k'} = \sum_{l=0}^{2(N-1)} l \cdot m_{x+y}^{k,k'} [I](l) \quad \text{Où } l = 0, 1, \dots, 2(N-1)$$

8/ Variance des sommes :

$$f_8^{k,k'} = \sum_{l=0}^{2(N-1)} (l - f_7)^2 m_{x+y}^{k,k'} [I](l)$$

9/ Entropie des sommes :

$$f_9^{k,k'} = - \sum_{l=0}^{2(N-1)} m_{x+y}^{k,k'} [I](l) \log \{m_{x+y}^{k,k'} [I](l)\}$$

10/ Variance des différences :

$$f_{10}^{k,k'} = \sum_{l=0}^{N-1} (l - f_{11})^2 m_{x-y}^{k,k'} [I](l)$$

11/ Entropie des différences :

$$f_{11}^{k,k'} = - \sum_{l=0}^{N-1} m_{x-y}^{k,k'} [I](l) \log \{ m_{x-y}^{k,k'} [I](l) \}$$

12/ Information sur la corrélation :

$$f_{12}^{k,k'} = \frac{f_9^{k,k'} - HXY1}{\max\{HX, HY\}}$$

Où  $HXY1 = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m^{k,k'} [I](i, j) \log \{ m_x^{k,k'} [I](i) m_y^{k,k'} [I](j) \}$

$$HX = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m_x^{k,k'} [I](i) \log \{ m_x^{k,k'} [I](i) \}$$

$$HY = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m_y^{k,k'} [I](j) \log \{ m_y^{k,k'} [I](j) \}$$

13/ Information sur la corrélation :

$$f_{13}^{k,k'} = \sqrt{(1 - \exp[-2.0(HXY2 - f_9)])}$$

Où  $HXY2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} m_x^{k,k'} [I](i) m_y^{k,k'} [I](j) \log \{ m_x^{k,k'} [I](i) m_y^{k,k'} [I](j) \}$

14/ Coefficient de corrélation maximal :

$$f_{14}^{k,k'} = (1 \text{ere plus grande valeur propre de } Q)^{\frac{1}{2}}$$

Où  $Q(i, j) = \sum_{k=0}^{N-1} \frac{m^{k,k'} [I](i, k) m^{k,k'} [I](j, k)}{m_x^{k,k'} [I](i) m_y^{k,k'} [I](k)}$

De manière à exploiter plus aisément les informations contenues dans ces matrices, nous exposons les indices les plus fréquemment utilisés comme : contraste, entropie, énergie, variance, corrélation, homogénéité

### 3.2.1.3 Choix de la méthode

Cette méthode constitue la référence en analyse de texture et sert souvent de méthode comparative pour les nouvelles approches.

La méthode des matrices de cooccurrences est très courante en analyse de texture, de nombreux articles De nombreux articles utilisent les matrices de cooccurrences pour identifier et classer les textures. [14]

Les paramètres issus des matrices de cooccurrences sont les plus utilisés en analyse de texture, notamment en analyse d'images médicales. [6]

Ainsi [6] déclare dans son étude dans le domaine d'imagerie médicales qu'il a obtenu des meilleurs résultats avec les matrices de cooccurrences par rapport à d'autres méthodes comme les filtres de Gabor, les ondelettes ou encore les statistiques d'ordre 1

Sans oublier qu'avec les statistiques du 1<sup>er</sup> ordre il n'y a pas d'informations sur la localisation du pixel. Il est donc nécessaire d'utiliser des méthodes d'ordre supérieur pour une analyse plus précise.

La méthode d'extraction de ces paramètres basée sur le calcul des matrices de cooccurrence est l'une des méthodes les plus proches de la notion de texture. [8]

Mais malgré ces avantages, les MC ont l'inconvénient d'être très coûteuse en temps de calcul. [14] [6]

### 3.2.2 Méthode d'Analyse Fractale

Grace au mathématicien Français Benoit Mandelbrot qui en fait l'objet d'une nouvelle discipline mathématique : La géométrie Fractale. En effet la nature n'est pas analytique, lisse, dérivable, mais elle est plutôt fractale. La géométrie fractale prend donc le relais et permet d'étudier avec succès, ces objets appelés objets fractals.

Le mathématicien Benoit Mandelbrot définissait l'adjectif « fractal » pour désigner des objets dont la géométrie complexe ne peut plus être caractérisée par une dimension entière. Ce phénomène est souvent exprimé sous forme de lois d'échelles statistiques spatiales ou temporelles et se caractérise donc principalement par des lois de puissances sur le comportement du système physique observé.

Ce concept, fréquemment rencontré dans différents domaines comme la géophysique, la biologie permet de donner une interprétation géométrique simple. Mandelbrot a introduit, à cet effet, la notion d'ensemble fractal (Mandelbrot, 1977), permettant de rendre compte du degré de régularité des organisations structurelles, liées au comportement du système physique.

La géométrie fractale est largement utilisée dans les problèmes d'analyses d'images en général et notamment dans le domaine médical, où elle trouve différentes applications et fournit divers résultats. Cependant, aucun état de l'art présentant les méthodes de cette géométrie et leurs applications n'existait dans la littérature

#### **Objet fractal ou non fractal ?**

Un objet est dit non fractal s'il n'y a pas d'apparition de nouvelles formes chaque fois qu'on zoome une de ses parties. Or dans le cas d'un objet fractal une nouvelle forme est apparue à chaque fois qu'une partie de l'objet est zoomé. Cette forme est plus au moins similaire à la totalité de l'objet lui-même. (Voir figure 3.1)

Ce qui nous vient maintenant à la tête une expression ou bien une explication concernant le choix de la méthode fractale : « La notion fractale est faite pour désigner des objets qui sont géométriquement compliquée (qui ont une forme irrégulière, très interrompue) » [8]

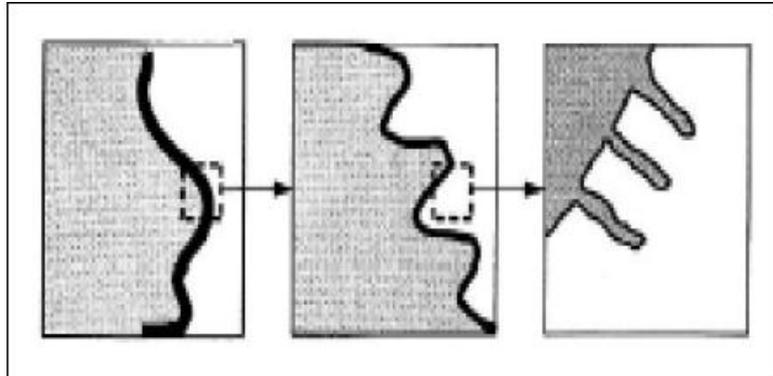


Figure 3.1 Objet Fractale

Ces dix dernières années, ces deux analyses (fractale et multi fractale) ont commencé à être utilisées en analyse d'images. La principale supposition fut qu'au lieu d'analyser un objet pour savoir s'il était fractal, on lui appliquait des traitements fractals indépendamment de sa possible invariance d'échelle.

Ce concept nouveau a fourni de plus en plus d'applications intéressantes dans divers domaines, notamment l'imagerie médicale.

### 3.2.2.1 Self similarité ou auto similarité dans les images

La propriété particulière d'une fractale dite autosimilarité (ou encore fractale à homothétie interne) est que chacune de ses parties reproduit sa totalité.

Et puisque on est dans le concept d'analyse fractale, il est indispensable de parler sur la notion de la dimension fractale, indice de référence de cette analyse.

### 3.2.2.2 Notion de la Dimension Fractale

Le terme de Dimension fractale (DF) a été introduit par *B.Mandelbrot* en 1970, la DF est avant tout un paramètre permettant de quantifier la complexité d'une image (quantifier le degré d'irrégularité et de fragmentation d'un ensemble géométrique). On peut par exemple utiliser ce paramètre pour classer des images en fonctions de leurs complexités. [8]

Plaçons-nous maintenant avec notre méthode d'analyse fractale une méthode proposée par [Alceu Ferraz Costa, Gabriel Humpire-Mamani, Agma Juci Machado Traina, Department of Computer Science, University of São Paulo, USP], qui s'appelle « Segmentation-based Fractal Texture Analysis », ou SFTA

### 3.2.2.3 Un algorithme efficace pour l'analyse fractale des textures : SFTA

L'algorithme d'extraction consiste à décomposer l'image d'entrée (image en niveau de gris) en un ensemble d'images binaires à partir de laquelle les dimensions fractales des frontières des régions sont calculées pour décrire les caractéristiques de texture segmentée. Pour cela un nouvel algorithme, appelé Two-Threshold Binary Decomposition (TTBD) est également proposé. [22]

#### Algorithme SFTA

Algorithm 1 SFTA extraction algorithm.

Require : Grayscale image  $I$  and number of thresholds  $nt$ .

Ensure : Feature vector VSFTA.

```

1:    $T \leftarrow \text{MultiLevelOtsu}(I; nt)$ 
2:    $T_A \leftarrow \{\{t_i, t_{i+1}\}: t_i, t_{i+1} \in T, i \in [1..|T| - 1]\}$ 
3:    $T_B \leftarrow \{\{t_i, n_l\}: t_i \in T, i \in [1..|T|]\}$ 
4:    $i \leftarrow 0$ 
5:   for  $\{\{t_\ell, t_u\}: \{t_\ell, t_u\} \in T_A \cup T_B\}$  do
6:      $I_b \leftarrow \text{TwoThresholdSegmentation}(I, t_\ell, t_u)$ 
7:      $\Delta(x, y) \leftarrow \text{FindBorders}(I_b)$ 
8:      $V_{SFTA}[i] \leftarrow \text{BoxCounting}(\Delta)$ 
9:      $V_{SFTA}[i + 1] \leftarrow \text{MeanGrayLevel}(I, I_b)$ 
10:     $V_{SFTA}[i + 2] \leftarrow \text{PixelCount}(I_b)$ 
11:     $i \leftarrow i + 3$ 
12:  end for
13:  return  $V_{SFTA}$ 

```

### 3.2.2.4 Choix de la méthode « SFTA »

On remarquant les résultats de cette technique présentée dans [22] par rapport à d'autres approches citée précédemment, nous remarquons que cet algorithme est plus performant et offre des bons résultats par exemple :

L'algorithme SFTA montre qu'il est 3,7 fois plus rapide que le filtre de Gabor

Et 1,6 fois plus rapide que l'approche de *Haralick*

Donc les résultats expérimentaux présentés dans cette étude [22] ont montré que la performance de SFTA est vraiment supérieure à celle de filtre de Gabor et Haralick, avec une grande précision et exactitude

Maintenant, on va y aller plus loin de l'analyse de texture, une phase de classification a été proposée pour profiter de ces vecteurs caractéristiques et ces bases de données, afin d'éteindre notre but concernant la classification des tumeurs cérébrales.

### **3.3 Classification avec les forêts aléatoires**

Les forêts aléatoires sont une méthode statistique non-paramétrique aux performances exceptionnelles. Elles ont été introduites par Breiman (2001) et sont de plus en plus utilisées pour traiter de nombreux et divers jeux de données.

Dans cette section, nous introduisons le contexte général dans lequel s'inscrivent les forêts aléatoires, les fondements mathématiques, ainsi que la famille de méthodes dont elles font partie.

#### **3.3.1 L'apprentissage statistique**

Le cadre mathématique de l'apprentissage statistique est le suivant. Soit  $\mathcal{L}_n = \{(X_1; Y_1), \dots, (X_n; Y_n)\}$  un échantillon d'apprentissage, c'est-à-dire une suite de vecteurs aléatoires indépendants et identiquement distribués, de même loi qu'un vecteur aléatoire  $(X, Y)$ . Le vecteur  $(X, Y)$  est indépendant de  $\mathcal{L}_n$  et sa loi est inconnue. L'entier naturel  $n$  désigne le nombre d'observations de l'échantillon d'apprentissage. [23]

Avant de décrire en détails la méthode des forêts aléatoires, nous présentons une famille plus large de méthodes statistiques à laquelle elle appartient : les méthodes d'ensemble.

#### **3.3.2 Les méthodes d'ensemble**

En quelques mots, Le principe général des méthodes d'ensemble (voir, Dietterich (2000)) est de construire une collection de prédicteurs, pour ensuite agréger l'ensemble de leurs prédictions. Dans un cadre de classification, l'agrégation revient à faire un vote majoritaire parmi les classes fournies par les prédicteurs.

Ce qu'il ne faut rater comme information, est que les méthodes d'ensemble (Bagging, Boosting, et Randomizing Outputs...) ont toutes un principe général commun, il s'agit de partir d'une règle de prédiction de base, puis de perturber cette règle de base.

On construit alors une collection de prédicteurs issus de différentes perturbations de la règle de base. Enfin on agrège l'ensemble des prédicteurs obtenus.

Pour chacune de ces méthodes, les auteurs montrent sur des simulations que le prédicteur agrégé final fait systématiquement mieux (en termes d'erreur de généralisation) que la règle de prédiction de base. Donc, en pratique, il apparaît que "perturber puis agréger" améliore les performances d'une méthode de prédiction donnée.

Maintenant on va y aller plus profond pour comprendre une des premières méthodes d'ensemble qui entre dans notre cadre d'étude sur la classification avec les forêts aléatoires : le Bagging

### 3.3.2.1 Le Bagging

Principe générale : agréger une collection de classifieurs faibles pour obtenir un meilleur classifieur. En général pour la classification, agrégation par vote majoritaire.

La méthode du Bagging a été introduite par Breiman (1996). Le mot Bagging est la contraction des mots Bootstrap et Aggregating. Le Bagging repose sur la méthode Bootstrap.

1 : Bootstrap : On construit  $q$  échantillons  $\mathcal{L}_n^{(1)}, \dots, \mathcal{L}_n^{(q)}$  à partir d'un seul échantillon de départ  $\mathcal{L}_n = (X_i, Y_i)_{i=1, \dots, n}$ .

- Un échantillon bootstrap  $\mathcal{L}_n^{(1)}$  est obtenu par tirage avec remise de  $n$  éléments parmi  $\mathcal{L}_n$  ou chaque observation  $(X_i, Y_i)$  a une probabilité  $\frac{1}{n}$  d'être tiré à chaque tirage.

2 : Chaque échantillon bootstrap  $\mathcal{L}_n^{(l)}$  sert à construire un classifieur  $\hat{g}_l, \forall l = 1, \dots, q$

3 : On agrège cette collection de classifieur  $\hat{g}_1, \dots, \hat{g}_q$ :

$$\hat{g}_{bag}(x) = \underset{K \in \{1, \dots, k\}}{\arg \max} \sum_{l=1}^q \mathbb{1}_{\hat{g}_l(x)=K}$$

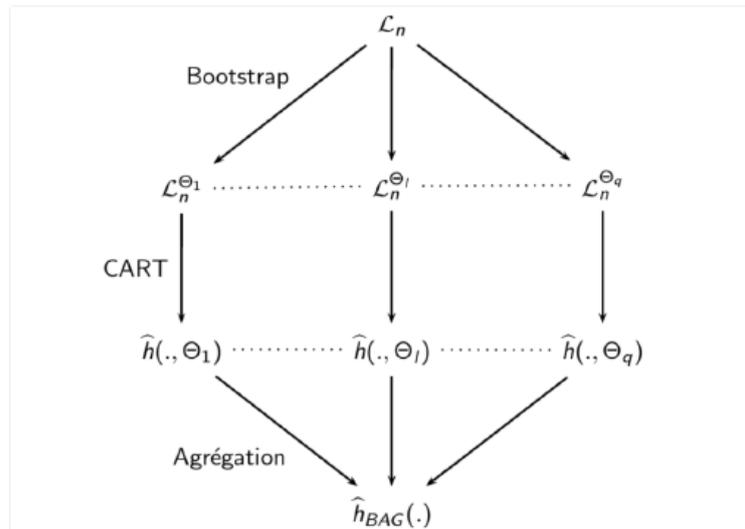


Figure 3.2 : Principe du Bagging

Notations :  $\theta_l, l = 1, \dots, q$  sont des variables indépendantes pour l'aléa de Bootstrap,  $\hat{h}(\cdot, \theta_l)$  est le prédicteur construit à partir de l'échantillon  $\mathcal{L}_n^{\theta_l}$ .

Bagging : Remarques

- Les classifieurs de la collection  $\hat{g}_1, \dots, \hat{g}_q$  ne sont pas indépendants car les échantillons ne le sont pas. Cependant le Bagging agit comme un réducteur de variance.
- Bagging = **Bootstrap Aggregating** : Principe non limité aux arbres de classification mais CART est le classifieur de base le plus couramment utilisé.

Bagging : Limites

- Les classifieurs de la collection  $\hat{g}_1, \dots, \hat{g}_q$  ne sont pas indépendants et ils sont même parfois **trop corrélés**
  - L'instabilité ne suffit pas à créer une collection d'arbres suffisamment diverse

Après ce passage qui parle sur le Bagging, en va s'intéresser maintenant à notre méthode de classification les forêts aléatoires qui sont en fait un cas particulier du Bagging.

### 3.3.2.1.1 Forêts Aléatoires

Les Forêts Aléatoires ont été introduites par Breiman (2001). C'est une méthode d'ensemble appliquée et limitée aux arbres CART.

Principe : décorrélérer la collection d'arbres provenant du Bagging en introduisant de l'aléa dans la construction de l'arbre.

- Plusieurs sources d'aléa ont été testées mais la méthode de Breiman (RF-RI) s'est imposée comme la méthode RF (Random Forest) par excellence.

Voici la définition générale des forêts aléatoires, donnée dans l'article de Breiman :

**Définition 1.** Soit  $\{\hat{h}(\cdot, \theta_1), \dots, \hat{h}(\cdot, \theta_q)\}$  une collection de prédicteurs par arbre, où  $(\theta_1, \dots, \theta_q)$  est une suite de variables aléatoires, c'est-à-dire, indépendante de l'échantillon d'apprentissage  $\mathcal{L}_n$ . Le prédicteur des forêts aléatoires est obtenu par agrégation de cette collection de prédicteurs.

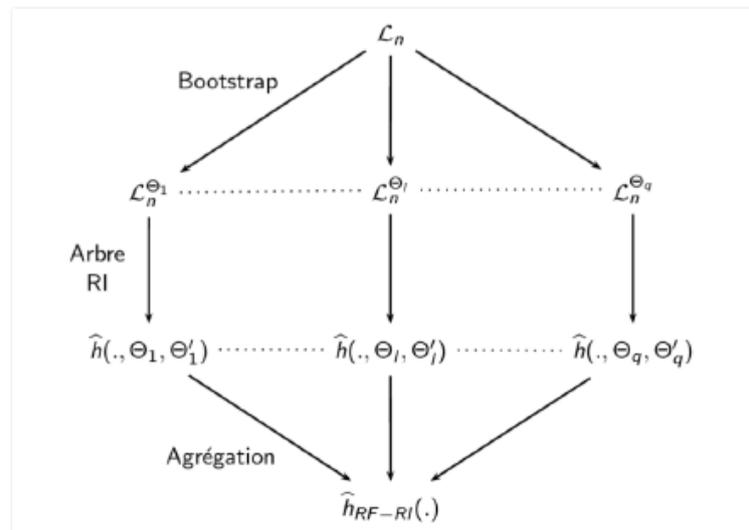


Figure 3.3 : Principe des Forêts Aléatoires

Notations :  $\theta_l, l = 1, \dots, q$  sont des variables indépendantes pour l'aléa de bootstrap,  $\theta'_l, l = 1, \dots, q$  sont des variables indépendantes pour l'aléa du prédicteur,  $\hat{h}(\cdot, \theta_l, \theta'_l)$  est le prédicteur aléatoire construit à partir de l'échantillon  $\mathcal{L}_n^{\theta_l}$ .

## Forêts Aléatoires : Méthode de Breiman

1. On tire  $q$  échantillons bootstrap  $\mathcal{L}_n^{(1)}, \dots, \mathcal{L}_n^{(q)}$  (comme pour le Bagging)
2. Sur chaque  $\mathcal{L}_n^{(l)}$ , on applique une variante de CART appelée RI (**R**andom **I**ntput) qui diffère à deux niveaux :
  - 2.1 Pour découper un nœud de l'arbre, on optimise la coupure sur un ensemble aléatoire de  $m \leq p$  variables (tirage sans remise de  $m$  parmi  $p$  variables)
  - 2.2 Les arbres ne sont pas élagués, on garde l'arbre maximal à chaque fois
3. Enfin, on agrège la collection de classifieurs obtenus  $\hat{g}_1, \dots, \hat{g}_q$

$$\hat{g}_{RF}(x) = \underset{K \in \{1, \dots, k\}}{\arg \max} \sum_{l=1}^q \mathbb{1}_{\hat{g}_l(x)=K}$$

On s'entend souvent de ce qu'on appelle l'erreur OOB, alors que ce qu'un OOB ?

## Forêts Aléatoires : erreur OOB

Taux d'erreur Out Of Bag : estimation acceptable du taux d'erreur théorique obtenue grâce au bootstrap.

Principe :

- Pour  $i = 1, \dots, n$  : On considère l'observation  $(X_i, Y_i) \in \mathcal{L}_n$ 
  1. on considère l'ensemble des échantillons bootstrap ne contenant pas  $(X_i, Y_i)$  (pour lesquels  $(X_i, Y_i)$  est OOB = "en dehors du bootstrap") ainsi que tous les arbres associés à ces échantillons.
  2. On prédit  $\hat{Y}_i$  en fonction de  $\hat{X}_i$  en agrégeant uniquement ces arbres
- On obtient alors le taux d'erreur OOB : En classification

En classification :

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\hat{Y}_i \neq Y_i}$$

Et parce que les Forêts Aléatoires est une méthode de type CART (les arbres CART) largement utilisée dans les méthodes d'ensemble.

### 3.3.3 Classification and regression trees CART

L'acronyme CART signifie Classification And Regression Trees. Il désigne une méthode statistique, introduite par Breiman et al. (1984). Designer pour la construction de modèles de prédiction (prédicteurs) à partir des données, ces modèles sont obtenus grâce à des partitions de données et puis chaque partie nous donne un modèle, et puis la combinaison de ces modèles nous donne un modèle final.

### 3.3.4 Choix de la méthode

Le choix de la méthode n'a pas été effectué au hasard, mais parce nous avons basés sur des critères qui sont adaptés pour notre étude de classification des tumeurs cérébrales. On cite parmi lesquels : une reconnaissance très rapide, Classification multi-classe par nature (high grade tumor Or low grade tumor), une méthode particulièrement bien adaptée pour traiter des données de très haute dimension telles que les textes et les images comme il a montré [23]. De plus en plus utilisées pour traiter de nombreux et divers jeux de données réelles. Les domaines d'application sont nombreux, citons par exemple l'étude des biopuces (c.f. Díaz- Uriarte and Alvarez de Andrés (2006)), l'écologie (c.f. Prasad et al. (2006)) ou encore la génomique (c.f. Goldstein et al. (2010)).

## 3.4 Conclusion

Dans ce chapitre, nous avons tenté de présenter nos différentes méthodes d'extraction d'attribut de texture des images IRM, nous avons vu les Matrices de Cooccurrences qui construisent une référence en Analyse de texture et la deuxième méthode, l'Analyse Fractale qui fait partie des méthodes les plus utilisées en raison de bonnes résultats obtenue. Ensuite nous avons donné une vision générale et purement mathématique sur notre méthode de classification les Forêts Aléatoires.

Dans le chapitre suivant, nous entamerons la partie pratique de ce mémoire en commençant par la phase de prétraitement de notre base de données, ensuite les résultats de notre travail avec discussion, en termine avec notre application développée avec Java FX qui est le fruit de cette thèse.

# **CHAPITRE 4**

## **CLASSIFICATION AVEC RANDOM FOREST APPLICATION AUX IMAGES IRM CEREBRALE**

## 4.1 Introduction

Ce chapitre présente deux travaux de nature appliquée. Dans la première section, nous traitons notre méthodologie de travail dès le prétraitement d'images de la base d'images (BRATS, 2015), puis la phase de la construction de la base de données et en termine cette section par illustrations des différents résultats obtenue dans la phase de la classification avec notre classifieur RF. Le but est de construire un modèle qui réussit à prédire la catégorie de la tumeur.

Dans la deuxième section du chapitre, nous présentons l'interface graphique de notre application développée avec Java FX, qui communique avec le logiciel Matlab, en vous présentons les différentes fonctionnalités. Afin de présenter à la fin le résultat obtenu.

## 4.2 Prétraitement d'images

Nous allons vous présenter dans ce contexte toutes les opérations effectuées dans notre travail, concernant, la préparation des images à partir de la base d'images IRM (BRATS, 2015) qui seront ensuite utiliser pour la construction des bases des données.

### 4.2.1 Base de données BRATS, 2015

BRATS, est l'acronyme de « Multimodal Brain Tumor Image Segmentation Benchmark », organisé en collaboration avec MICCAI 2012,2013 [9]

Ensemble d'images IRM, en effet ce sont des images cervicale de 65 patients qui souffrent de la maladie *Gliome*.

Les données d'image BRATS et annotations manuelles continuent d'être accessibles au public par le biais d'un système d'évaluation en ligne comme une ressource d'analyse comparative en cours.

#### **Définition : Glioma**

Glioma, est une tumeur cérébrale primaire considéré comme le plus fréquent chez les adultes, provenant probablement de cellules gliales (localisée dans le système nerveux).

Après ces brèves définitions sur la base de d'images BRATS, 2015 et la maladie de Gliome, en passe vers la pratique : Conversion des images IRM

## 4.2.2 Conversion des images

Premièrement, la conversion des images est plus que nécessaire, Pourquoi ? Car notre base d'images (BRATS, 2015) nous proposons des images sous l'extension (.mha). Et pour cela il faut convertir ces images en autre extension comme par exemple (.nii) à l'aide d'un logiciel open source qui s'appelle « *Slicer* » (Voir Figure XX). Ce logiciel offre une interface dynamique afin de pouvoir afficher l'image sélectionnée de la base d'images et observer la lésion avec différentes coupes anatomiques (Axial, Sagittal, Coronal) et puis l'enregistrer en (.nii) dans un fichier à part.

Deuxièmement, grâce au logiciel aussi open source « *ImageJ* » nous pouvons ouvrir ces d'images (enregistrée en format .nii) afin de choisir la coupe à laquelle nous voyons qu'elle montre clairement la lésion tumorale, et puis l'enregistrer en format (.Jpg) qui nous permet de charger finalement cette image dans un programme de caractérisation, on parle des programmes d'Analyse Fractale ou bien Les matrices des Cooccurrences avec *Matlab*.

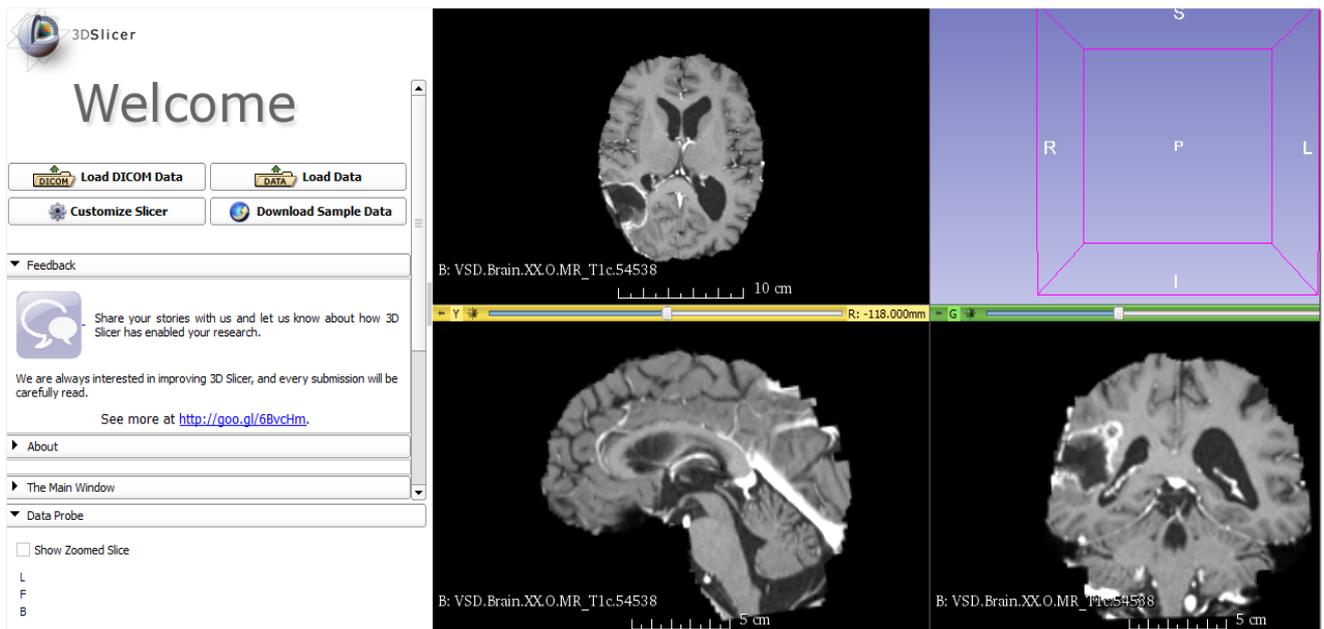


Figure 4.1 Interface graphique du logiciel *Slicer*



Figure 4.2 : interface graphique du logiciel *ImageJ*

En raison de bien organiser les choses, nous avons rassemblé toutes les images ou bien toutes les coupes dans deux fichiers différents, un pour le Haut grade et l'autre pour Bas grade pour faciliter la tâche suivante.

**Remarque :**

Cette dernière opération, Capture d'une coupe d'image a été effectuée après une consultation avec un médecin spécialiste en Radiologie (ex Radiologue en CHU Batna), qui nous a dit « ce qui nous intéresse est toute la lésion et non pas seulement le contour blanc ». (Figure 4.2).

### 4.3 Caractérisation et construction des bases de données

Comme nous avons dit précédemment, nous avons travaillé avec deux techniques d'analyse de texture pour construire nos bases de données avec Matlab.

Un processus a été effectué pour les deux états (Haut grade, Bas grade), c'est-à-dire qu'on va construire des bases de données avec les deux techniques de caractérisation pour le Haut grade et la même chose pour le Bas grade. Pourquoi ? C'est juste pour bien organiser les choses.

Mais en réalité une base de données est un mélange entre le high grade et le Low grade pour une seule technique, le schéma suivant va vous montrer un petit peu l'organigramme de construction des bases de données.

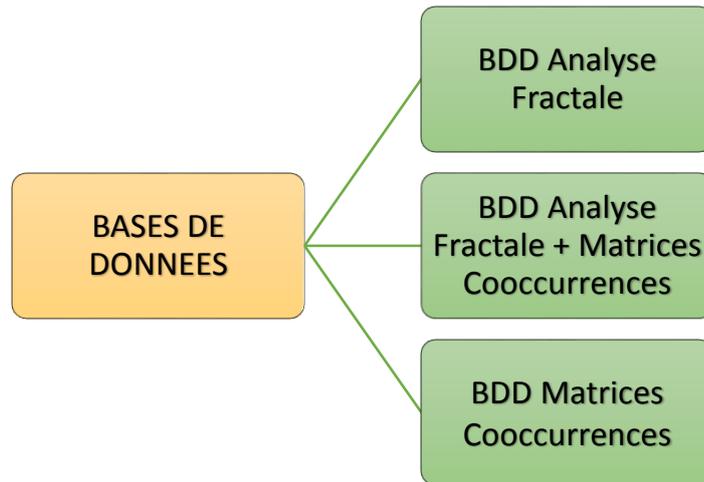


Figure 4.3 : les différentes bases de données

Attendez ce n'est pas encore terminé ! Avec chaque technique de caractérisation, nous allons créer des bases de données et non pas une seule base de données.

Alors pourquoi ne pas créer une base de données pour chaque technique ? Au fait ce que nous voulons dire, est qu'il y a une relation entre ces bases de données et le taux de classification ?

En raison de bien exploiter ces techniques de caractérisation (Analyse Fractale et Matrices Cooccurrences) avec notre technique de classification, nous nous sommes dits pourquoi ne pas créer des bases de données en termes de taille de la fenêtre, c'est-à-dire que nous allons exécuter le programme de caractérisation (par exemple : l'analyse fractale) pour construire des bases de données mais à chaque fois nous changeons la taille de la fenêtre, autrement dit, nous voulons analyser la texture d'une image IRM.

Est-ce que la taille de la fenêtre a un impact sur le taux de classification ? On verra la réponse dans la partie où nous allons présenter nos résultats de la classification

Le schéma suivant va simplifier la vision, Par exemple *BDD Matrices Cooccurrences\_64*, veut dire que nous avons travaillé avec la technique des MC où la taille de la fenêtre est  $64*64$

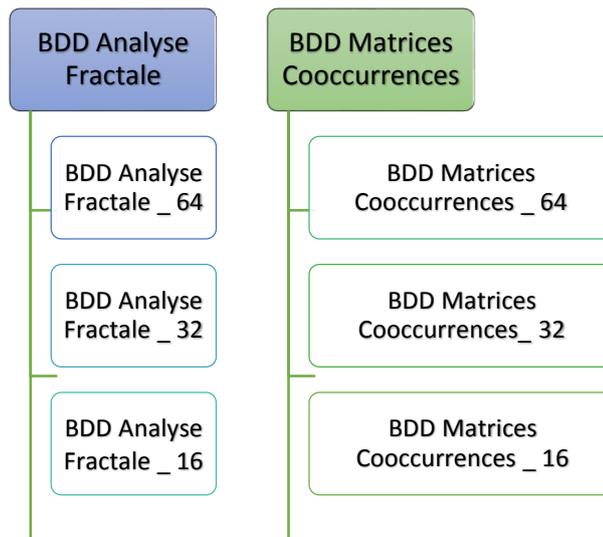


Figure 4.4 : les différentes tailles des fenêtres des bases de données

Un autre aspect aussi très important, est le nombre de paramètres, nous avons travaillé avec 48 attributs pour l'analyse fractale et 22 attributs pour les matrices de Cooccurrences, ensuite nous avons réduit le nombre de paramètres des matrices de Cooccurrences jusqu'aux Six attributs qui sont les plus utilisés avec les statistiques d'ordre deux (les indices de *Haralick*)

Est-ce que le nombre de paramètres a un impact aussi sur le taux de classification ? C'est ce que nous vous invitons à découvrir dans la partie résultat.

## 4.4 Résultat et Discussion

Pour cette première expérience, l'élaboration de la classification a été faite avec un nombre d'arbre égal à 500, et nombre d'itération égale à 1.

### 4.4.1 Pourquoi Nombre d'itération égale à 1 ?

Un autre paramètre : Nombre d'itérations, cela veut dire combien de fois nous voulons faire l'apprentissage. En raison des résultats obtenus avec différents nombre d'itérations. Nous avons remarqué que le changement de TC est insignifiant ou bien négligeable, c'est pour cela nous avons continué à utiliser un nombre d'itérations égale à 1.

#### 4.4.2 Pourquoi Nombre d'arbre égal à 500 ?

En effet, d'après nos expérimentations avec ce programme (RF) dans Matlab, nous avons obtenu des bons résultats en précisant le nombre d'arbres à 500 contrairement avec l'utilisation d'autres paramètres, un graphe justificatif parle bien sur ce point (Figure 4.5).

La deuxième chose, est que d'après les travaux précédents présents dans la littérature, beaucoup d'eux ont utilisés le RF avec un nombre d'arbres égal à 500 où ils ont obtenu aussi des bons résultats [23].

Commençant par un graphe justificatif qui représente le taux de classification (TC) en fonction du nombre d'arbre en utilisant le RF avec la BDD (AF+MC, taille 64\*64) et voilà le résultat :

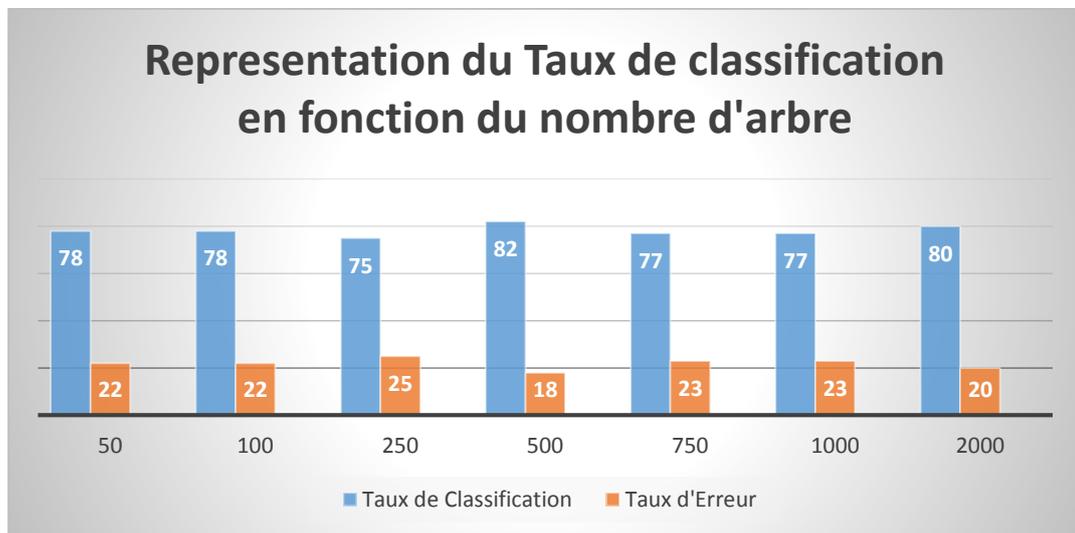


Figure 4.5 : représentation du Taux de classification en fonction du nombre d'arbre

Maintenant, le résultat confirme notre choix concernant le nombre d'arbre égal à 500, l'obtention d'un TC = 82% est le meilleur résultat. Et pour la confirmer, nous allons tester encore une fois ces paramètres mais avec un programme de RF en utilisant une autre BDD, par exemple dans le cas suivant nous allons utiliser la BDD (AF, taille 64\*64)

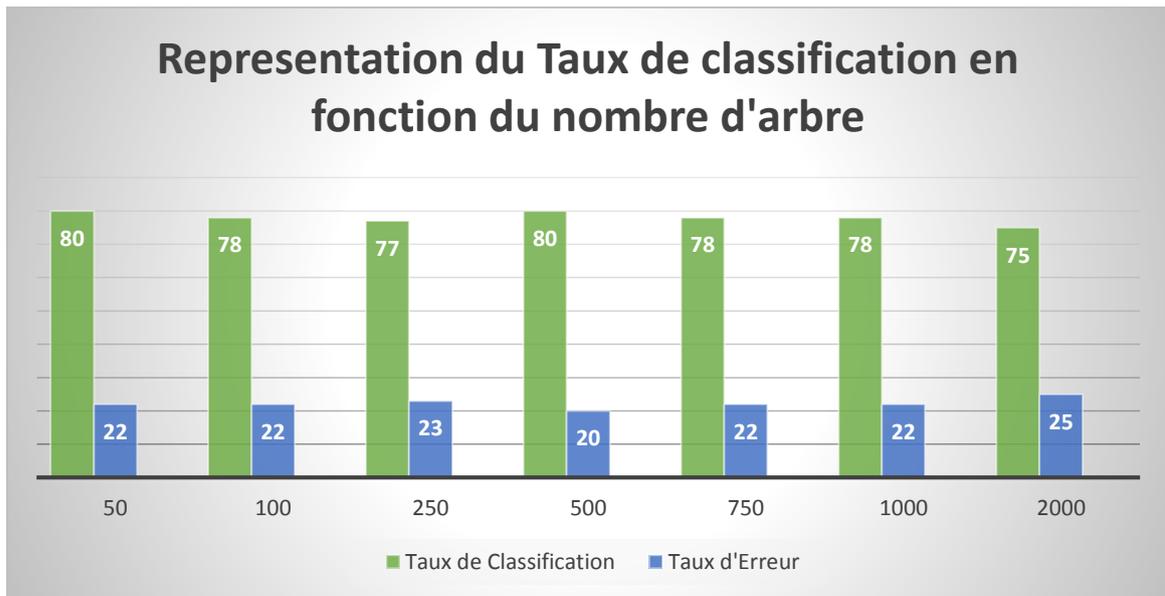


Figure 4.6 : représentation du Taux de classification en fonction du nombre d'arbre (2)

#### 4.4.3 Evaluations de la performance de classification

Plusieurs classifications ont été effectuées avec différents attributs :

1. Evaluer la performance de chaque technique avec différentes fenêtres

- {RF (AF 16 \* 16)} {RF (AF 32 \* 32)} {RF (AF 32 \* 32)}
- {RF (MC 16 \* 16)} {RF (MC 32 \* 32)} {RF (MC 64 \* 64)}
- {RF (MC\* 16 \* 16)} {RF (MC\* 32 \* 32)} {RF (MC\* 64 \* 64)}

2. Evaluer la performance, en utilisant la combinaison des deux techniques précédentes

{RF (AF + MC 16 \* 16)} {RF (AF + MC 32 \* 32)} {RF (AF + MC 64 \* 64)}

{RF (AF + MC\* 16 \* 16)} {RF (AF + MC\* 32 \* 32)} {RF (AF + MC\* 64 \* 64)}

#### Remarque :

MC\* veut dire que nous avons utilisé la technique MC mais seulement avec 6 attributs (une sorte de sélection de variables les plus utilisées de *Haralick*)

### Fenêtre de taille 16\*16

Technique utilisée	Taux de classification	Sensibilité/Spécificité	
RF(AF)	TC = 63%	Sn= 0.56	Sp= 0.68
RF (MC)	TC = 60%	Sn= 0.54	Sp= 0.64
RF (MC*)	TC = 69%	Sn= 0.63	Sp= 0.72
RF (AF + MC)	TC = 71%	Sn= 0.75	Sp= 0.70
RF (AF + MC*)	TC = 73%	Sn= 0.67	Sp= 0.78

Tableau 4.1 : Résultats de la classification en utilisant la fenêtre 16\* 16

Nous remarquons dans ce tableau que la combinaison RF (AF + MC\*) a donné le meilleur taux de classification 73% par rapport aux autres attributs.

### Fenêtre de taille 32\*32

Technique utilisée	Taux de classification	Sensibilité/Spécificité	
RF(AF)	TC = 70%	Sn= 0.65	Sp= 0.74
RF (MC)	TC = 64%	Sn= 0.60	Sp= 0.68
RF (MC*)	TC = 67 %	Sn= 0.61	Sp= 0.70
RF (AF + MC)	TC = 74%	Sn= 0.75	Sp=0.73
RF (AF + MC*)	TC = 75%	Sn= 0.76	Sp= 0.75

Tableau 4.2 : Résultats de la classification en utilisant la fenêtre 32\* 32

Dans ce cas, les deux remarques principales sont :

Nous pouvons constater l'augmentation des Taux de classification avec les différentes techniques par rapport aux résultats précédents présentées dans le tableau [4.1]

La deuxième chose, c'est que la combinaison RF (AF + MC\*) est toujours à la hauteur pour l'instant avec un TC égale à 75% bien que RF (AF + MC) donne aussi un TC de 74%. Donc il n y a pas un grand changement.

## Fenêtre de taille 64\*64

Technique utilisée	Taux de classification	Sensibilité/Spécificité	
RF(AF)	TC = 80%	Sn= 0.73	Sp= 0.85
RF (MC)	TC = 73%	Sn= 0.66	Sp= 0.79
RF (MC*)	TC = 71%	Sn= 0.69	Sp= 0.72
RF (AF + MC)	TC = 82%	Sn= 0.82	Sp= 0.82
RF (AF + MC*)	TC = 78%	Sn= 0.71	Sp=0.79

Tableau 4.3 : Résultats de la classification en utilisant la fenêtre 64\* 64

Ici, comme vous voyez, le changement est remarquable

La première chose remarquable, est l'obtention d'un TC égale à 82%, le meilleur résultat par rapport aux différentes techniques testées avec différentes fenêtres.

Et pour bien éclairer les résultats de chaque technique, nous allons les présenter avec les différentes tailles de la fenêtre.

- **L'Analyse Fractale**

Le tableau suivant présente les TC obtenues avec les différentes tailles de fenêtre

TECHNIQUE UTILISEE:	TAILE DE FENETRE	TAUX CLASSIFICATION	SENSIBILITE/SPECIFICITE	
RF(AF)	<b>16*16</b>	TC = 63%	Sn= 0.56	Sp= 0.68
RF(AF)	<b>32*32</b>	TC = 70%	Sn= 0.65	Sp= 0.74
RF(AF)	<b>64*64</b>	TC = 80%	Sn= 0.73	Sp= 0.85

Tableau 4.4 : Résultats de la classification en utilisant la technique RF (AF)

Dans un premier temps, il faut dire que le meilleur résultat obtenu avec l'AF est celui où nous avons utilisé une fenêtre de taille 64\*64 avec un TC égale à 80% et une spécificité égale à 85% cela veut dire qu'on a obtenu un bon résultat. Ce qui confirme l'hypothèse « une fenêtre de taille 64\*64 est recommandée dans la technique de l'analyse fractale » [22]

La deuxième chose remarquable, c'est la relativité, nous pouvons distinguer que l'augmentation de la taille de la fenêtre est accompagnée avec l'augmentation de TC.

## Pourquoi ? Comment interpréter ce résultat ?

En effet, on va extraire une partie de la base de données de l'AF avec la fenêtre 16\*16

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
8.89338235	0.02941176	0.18501873	0.18501873	0.07993708	-0.054057	0.02941176	0.9353644	0.17724741	0.98529412	0.98529412	0.96691176
9.11397059	0.02941176	0.18501873	0.18501873	0.07993708	0.054057	0.02941176	0.9353644	0.17724741	0.98529412	0.98529412	0.96691176
4.08823529	0.05514706	0.20155573	0.20155573	0.15993723	0.07443213	0.05514706	0.87989295	0.32646812	0.97242647	0.97242647	0.9375

Figure 4.7 : presentation d'une partie de la base de donnees en utilisant la fenetre 16\* 16

Nous pouvons déduire maintenant pourquoi un TC avec une BDD 64\*64 est meilleur que celle avec une BDD de taille 16\*16. Ce que nous voulons dire est que pendant la construction de la BDD (base des vecteurs caractéristiques) avec une fenetre de taille 16\*16, nous amènent parfois à une situation non déterminée, avec Matlab nous remarquons la présence de NAN au lieu d'un vecteur de valeurs !

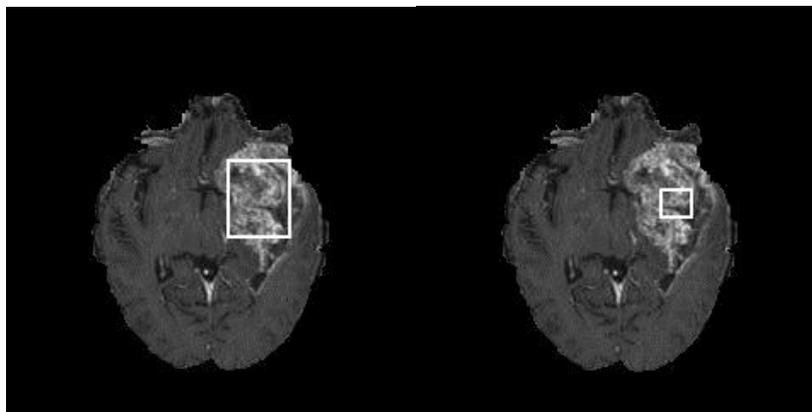


Figure 4.8 : la différence entre les tailles des fenetres

Cela veut dire que cette fenetre est vraiment petite pour caractériser toute la lésion tumorale présente dans l'image IRM. À cet effet, et avec la présence de quelques vecteurs dans la base de données qui sont égaux à zéros, contrairement à une base de données de taille 64\*64. Ce qui va influencer sur le taux de classification.

- **Matrices de Cooccurrences**

TECHNIQUES UTILISEES	TAILEE DE FENETRE	TAUX CLASSIFICATION	D SENSIBILITE/SPECIFICITE	
RF (MC)	<b>16*16</b>	TC = 60%	Sn= 0.54	Sp= 0.64
RF (MC)	<b>32*32</b>	TC = 64%	Sn= 0.60	Sp= 0.68
RF (MC)	<b>64*64</b>	TC = 73%	Sn= 0.66	Sp= 0.79

Tableau 4.5 : Résultats de la classification en utilisant la technique RF (MC)

Le Tableau suivant va vous présenter le TC obtenue avec MC\*

TECHNIQUES UTILISEES	TAILLEE DE FENETRE	TAUX CLASSIFICATION	D SENSIBILITE/SPECIFICITE	
RF (MC*)	<b>16*16</b>	TC = 69%	Sn= 0.63	Sp= 0.72
RF (MC*)	<b>32*32</b>	TC = 67 %	Sn= 0.61	Sp= 0.70
RF (MC*)	<b>64*64</b>	TC = 71%	Sn= 0.69	Sp= 0.72

Tableau 4.6 : Résultats de la classification en utilisant la technique RF (MC\*)

Ici, nous remarquons que la construction des bases de données de Matrices de Cooccurrences avec seulement 6 attributs a donné un taux de classification meilleur que celles des MC où nous avons travaillé avec les 22 attributs de Haralick. A part la dernière case, on parle dans le cas de la fenêtre taille 64\*64. Ou les résultats sont presque les mêmes, bien que MC donne mieux

Ce qui confirme le meilleur résultat obtenu dans notre approche RF (AF + MC)

Technique Utilisée		
Taille de la Fenêtre	MC	MC*
<b>16* 16</b>	60	69
<b>32*32</b>	64	67
<b>64*64</b>	73	71

Tableau 4.7 : comparaison des résultats entre MC et MC\*

- **Analyse fractale et Matrices de Cooccurrences**

Dans un premier temps, nous allons commencer par un tableau

TECHNIQUES UTILISEES	TAILLEE DE FENETRE	TAUX CLASSIFICATION	D SENSIBILITE/SPECIFICITE	
RF (AF + MC)	<b>16*16</b>	TC = 71%	Sn= 0.75	Sp= 0.70
RF (AF + MC)	<b>32*32</b>	TC = 74%	Sn= 0.75	Sp=0.73
RF (AF + MC)	<b>64*64</b>	TC = 82%	Sn= 0.82	Sp= 0.82

Tableau 4.8 : Résultats de la classification en utilisant la combinaison RF (AF+MC)

Et ensuite,

TECHNIQUES UTILISEES	TAILLEE DE FENETRE	TAUX CLASSIFICATION	D SENSIBILITE/SPECIFICITE	
RF (AF + MC *)	16*16	TC = 73%	Sn= 0.67	Sp= 0.78
RF (AF + MC *)	32*32	TC = 75%	Sn= 0.70	Sp= 0.75
RF (AF + MC *)	64*64	TC = 78%	Sn= 0.71	Sp=0.79

Tableau 4.9 : Résultats de la classification en utilisant la combinaison RF (AF+MC\*)

Globalement, vous pouvez remarquer que les TC obtenus avec les combinaisons RF (AF+MC), RF (AF+MC\*) sont un peu plus élevé par rapport aux TC obtenus avec l'utilisation de ces techniques chacune toute seul qui sont déjà présentés dans les tableaux précédents.

ET par ce que le graphe parle bien, nous avons choisi de vous présenter un graphe qui exprime le TC, la sensibilité et la spécificité, dans le cas d'une fenêtre 64\*64 par ce que maintenant nous savons que c'est le paramètre qui nous a donné des meilleurs résultats.

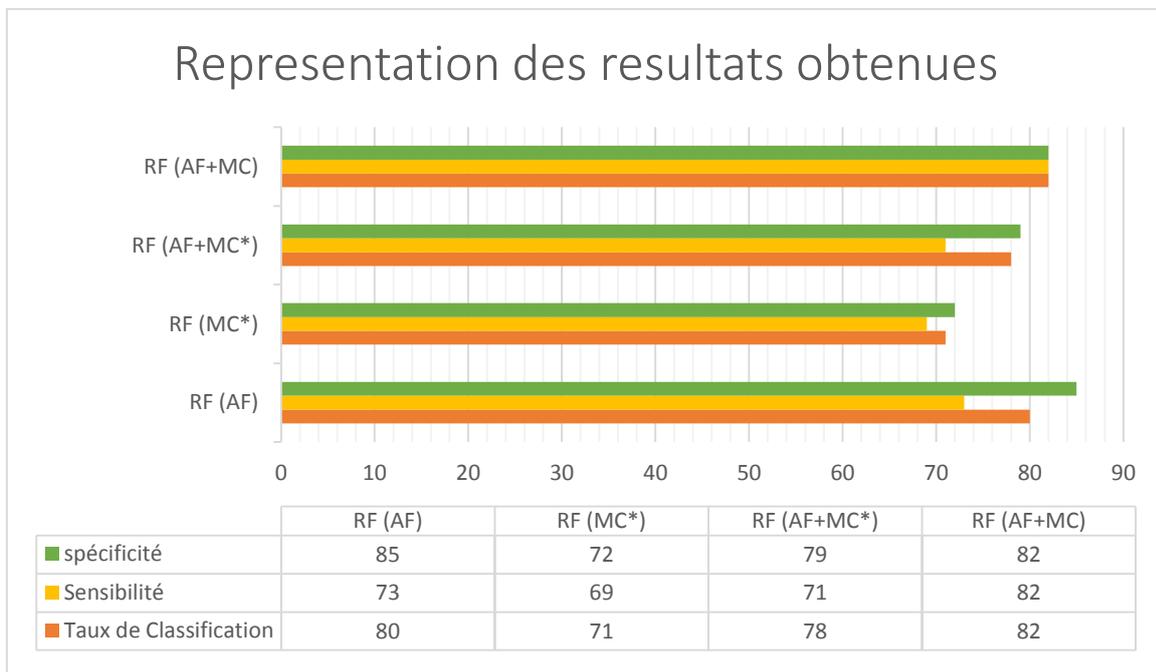


Figure 4.9 : Comparaison des meilleurs résultats obtenus

Nous pouvons dire selon le graphe que la technique qui a donné des bons résultats : RF (AF+MC) avec un taux de classification 0.8219. Donc un taux d'erreur de 0.1781%.

#### 4.4.4 Comment mesurer la performance du système ?

F-score : connu aussi sous le nom F-mesure, correspond à un compromis de la précision et de rappel donnant la performance du système. Ce compromis est donné de manière simple par la moyenne harmonique de la précision et de rappel

$$F - score = 2 \frac{P * R}{P + R}$$

Où

Précision : Proportion de solutions trouvées qui sont pertinentes. Mesure la capacité du système à refuser les solutions non-pertinentes. Donné par :

$$Precision : P = \frac{Vp}{Vp+Fp}$$

Rappel : Proportion des solutions pertinentes qui sont trouvées. Mesure la capacité du système à donner toutes les solutions pertinentes. Donné par :

$$Rappel : R = \frac{Vp}{Vp+Fn}$$

C'est quoi encore ces :  $Vp$ ,  $Fp$  et  $Fn$  ? On parle bien sûr de la sensibilité et la spécificité. Alors

La sensibilité (ou sélectivité) d'un test mesure sa capacité à donner un résultat positif lorsqu'une hypothèse est vérifiée. Elle s'oppose à la spécificité, qui mesure la capacité d'un test à donner un résultat négatif lorsque l'hypothèse n'est pas vérifiée.

$$Sensibilité = \frac{Vp}{Vp+Fn} \qquad Spécificité = \frac{Vn}{Vn+Fp}$$

Nous passons maintenant pour afficher les résultats ou bien les performances de notre modèle, nous avons choisi bien sûr le modèle qui nous a donné le meilleur résultat, on parle su RF (AF+MC 64\*64)

Pour la sensibilité et la spécificité, les résultats sont déjà affichés dans le tableau 4.8

Sensibilité = 0.82

Spécificité = 0.82

### La matrice de Confusion :

	P	N
P	VP : 4.62	FP : 0.74
N	FN : 0.88	VN : 0.21

Tableau 4.10 : La matrice de confusion de la technique RF (AF+MC)

Retournons à nos variables Précision et Rappel, d'après les calculs, on a obtenu comme :

Précision = 0.86

Rappel = 0.84

Donc maintenant nous pouvons calculer notre F-score

F-score = 0.85

### Remarque

- **La taille de la fenêtre**

Comme vous pouvez le constater selon les résultats présentés au-dessus, ce paramètre a un impact sur le TC, une classification avec RF en utilisant des bases de données construites avec des vecteurs caractéristiques obtenues lors de la caractérisation d'images IRM avec une fenêtre de 64\*64 est meilleur par rapport aux autres. Donc la taille de la fenêtre a un impact sur le TC.

- **Le nombre de variable**

On a choisi la technique des MC pour appliquer ce paramétrage, est finalement une classification avec seulement 6 attributs ce qui nous a donnée en globale des bons résultats par rapport à celle où nous avons utilisé tous les indices de *Haralick*. Ce qui confirme pourquoi ces paramètres sont les plus couramment rencontrées dans la littérature.

Maintenant on passe à la présentation de notre application qui va résumer tout ce processus.

## 4.5 Application Java FX

Notre application a été en réalité développée en utilisant deux environnements connus, on parle de NetBeans et Matlab, et avec différents langages de programmation comme Java, CSS et FXML.

### Pourquoi cette combinaison ?

Notre projet est constitué de différentes classes, comme vous pouvez le voir dans la figure ci-dessus, l'approche orientée objet se base sur l'utilisation de différentes classes partagées dans un seul package au contraire à l'approche classique de développement, cela va vraiment faciliter la tâche et lui rendre flexible et compréhensible.

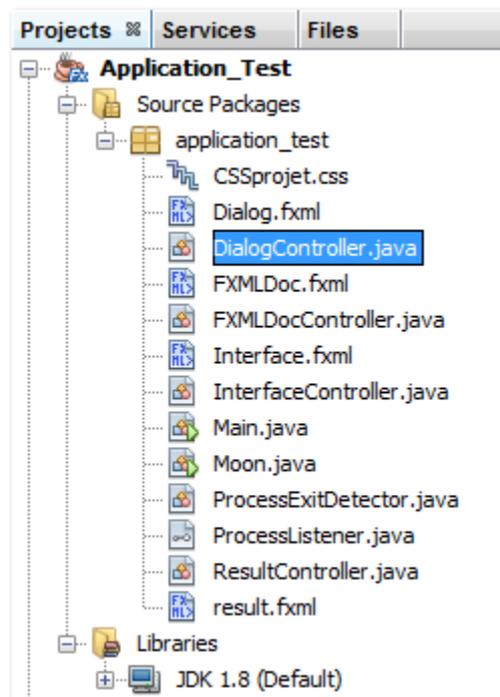


Figure 4.10 : présentation des classes du projet

En réalité ce n'est pas que du Java, mais nous avons utilisé un langage récent qui s'appelle JAVA FX

### C'est quoi JAVA FX ?

Avec l'apparition de Java 8 en mars 2014, Java FX devient l'outil de création d'interface graphique ('GUI toolkit') officiel du langage Java, pour toutes les sortes d'applications (applications mobiles, applications sur poste de travail, applications Web...). Java FX est un ensemble de packages graphiques et médias qui permet aux développeurs de concevoir, créer, tester, déboguer et déployer des applications client riche qui fonctionnent de façon uniforme dans les différentes plates-formes.

Vous remarquez peut être, l'absence de l'extension .javafx dans la figure précédente. La réponse, c'est que dès la création d'un projet Java FX en NetBeans, il vous donne l'accès de travailler avec une classe FXML, une autre Java, et une autre classe CSS pour le définir le style de l'application

### Alors c'est encore FXML ?

Un langage de balisage basé sur XML crée par Oracle pour définir l'interface utilisateur d'une application Java FX

Vous comprenez maintenant que notre interface a été designer en utilisant FXML, et parce que l'informatique facilite les choses, on a utilisé un outil de conception graphique : Java FX Scene Builder

### C'est quoi l'intérêt d'utiliser cet outil ?

Au lieu de programmer l'interface (définir la taille du bouton, définir sa place dans l'interface, programmer les panneaux...etc.) Java Fx scene builder facilite la tâche et vous permet de designer l'interface de manière dynamique

Et nous montrons maintenant un aperçu sur la fenêtre principale de notre application avec des brèves explications sur les différentes fonctionnalités

Load Image : nous permettons de charger l'image avec le nom du fichier qui va être affiché à coté comme vous vouez dans la (figure 4.11)



Figure 4.11 présentation de l'interface principale de notre application

Start Process : après le sélectionnent des pixels, ce bouton nous donne la main à choisir un mode d'exécution à travers une fenêtre (Figure 4.12)

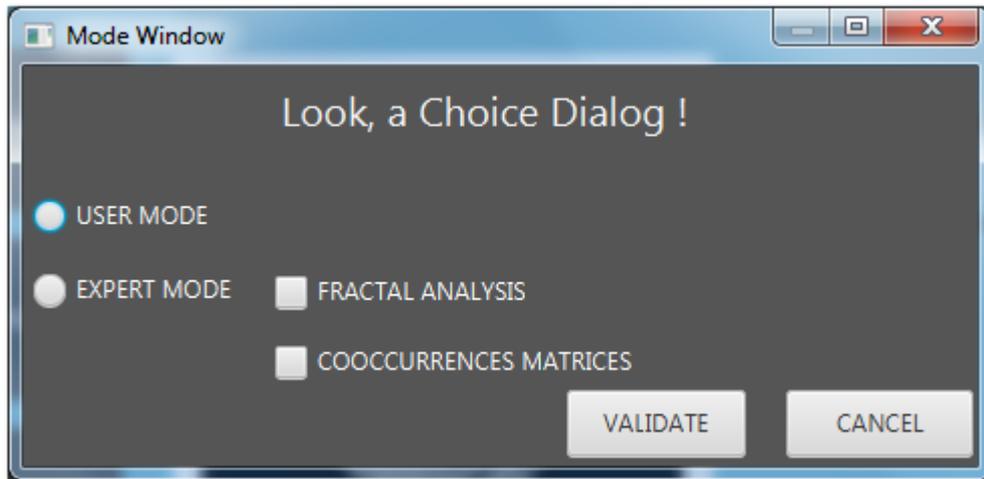


Figure 4.12 présentation de la fenêtre : Mode Window

Result : l'accès à ce bouton est interdit durant l'exécution du programme, et une fois le programme est terminé. Une fenêtre va être affichée pour nous donne le résultat final (figure 4.13)

Comme vous voyez le résultat est la catégorie de la tumeur, avec un taux de classification

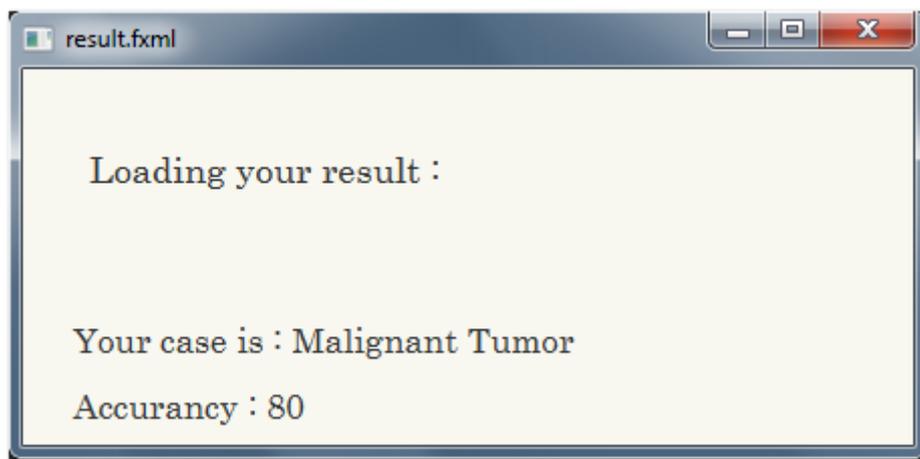


Figure 4.13 présentation de la fenêtre : Result Window

## 4.6 CONCLUSION

Au terme de ce chapitre, nous avons présenté la méthodologie utilisée dans ce travail, ensuite nous avons démontré que la classification des images texturées avec Random Forest basé sur une combinaison de caractéristiques d'Analyse fractales (AF) et Matrices de Cooccurrences (MC) a donné le meilleur taux de classification par rapport aux différents résultats obtenus.

Nous avons conclu que la dimension fractale a donné un bon résultat (80%) dans notre étude sur les images IRM, ce qui confirme que cette technique est conforme dans le cas des images de texturés irréguliers (Images IRM)

Ensuite, nous avons vu notre application, une interface graphique qui prend une image d'IRM cérébrale en entrée, afin de donner en sortie le résultat de la classification avec le taux de la classification.

## CONCLUSION GENERALE

Des millions d'images médicales sont produites chaque année dans le monde pour établir un diagnostic ou contrôler une action thérapeutique. Ces images proviennent principalement grâce aux modalités d'acquisition, tel que, l'IRM.

Les images médicales fournissent des informations sur la forme ou le fonctionnement des organes du corps humain. Malheureusement, ces informations sont extrêmement difficiles à exploiter de manière quantitative et objective.

L'étude que nous avons effectuée apporte des éléments de réponses à un problème de distinction entre une tumeur maligne ou bénigne. L'analyse automatique (semi-automatique) des images médicales et plus particulièrement l'analyse de texture peuvent offrir un ensemble de nouveaux outils d'aide au diagnostic. Parmi ceux-ci, l'extraction de paramètres quantitatifs objectif sur la texture.

La caractérisation des images médicales IRM, un concept très utilisé récemment dans divers domaines, notamment, l'imagerie médicale offre la possibilité d'extraire des informations pertinentes de la texture, afin de l'exploiter ensuite pour la distinction de la catégorie de la tumeur.

La classification joue un rôle très important, qui sert à nous donner une réponse claire sur la catégorie de la tumeur dans l'image IRM, avec Random Forest (RF) en utilisant deux techniques de caractérisation ou bien d'extraction d'attributs de textures, nous avons obtenu un taux de classification de 82%

La création de logiciels (application) dédiés à l'analyse d'images médicales doit permettre d'optimiser leur exploitation, pour le plus grand bénéfice du patient et du médecin. Un traitement d'image IRM en entrée qui donne une image en sortie avec le type de la tumeur c'est l'idéal pour les médecins spécialiste qui n'arrive pas cependant à donner un diagnostic fiable.

## PERSPECTIVES

De nombreuses extensions peuvent être envisagées et des améliorations peuvent également être apportées à ce travail.

1. L'utilisation de différentes modalités d'IRM au lieu d'utiliser qu'une seule (T1C), afin de combiner entre les différentes images IRM pondérées en (T1, T2, et Flair) en raison d'obtenir des bases de données variées qui contiennent une multitude d'informations, ce qui va aider les médecins spécialistes à donner un diagnostic avancé et plus précis.
2. Il serait intéressant d'utiliser des bases de données d'images médicales qui nous aident plus tard non seulement de faire la différence entre les catégories de la tumeur, mais aussi connaître le stade, est-ce que la tumeur est dans un état avancé ? et pourquoi définir la région ciblée dans le cerveau. C'est vraiment l'aide au diagnostic.
3. Un autre axe de recherche théorique serait d'essayer d'obtenir des résultats sur l'importance des variables. En effet, l'indice d'importance des variables des forêts aléatoires réussit à identifier les variables d'intérêt et les variables de bruit. Cependant, à ce jour, aucun résultat théorique n'explique ce bon comportement.

## Bibliographie

- [1] CHIKH.M.A, « Cours Apprentissage Artificiel », Université de Tlemcen, Faculté de Technologie
- [2] BENOIT Scherrer, « Segmentation des tissus ET structures sur les IRM cérébrales », Thèse de Doctorat, L'Institut Polytechnique de Grenoble, Décembre 2008
- [3] Les tumeurs du cerveau, collection Guides de référence Cancer info, INCa, juin 2010.
- [4] Les tumeurs cérébrales et du système nerveux central, Ligue suisse contre le cancer, Berne, 2009
- [5] KEDDOU Nadjat, ZEREG Mounia Feriel « Extraction des paramètres caractéristiques des images IRM pour la classification des tumeurs cérébrales », Thèse de Master, Université de Boumerdes, 2015
- [6] RENAUD Lopes, « Analyse Fractale et Multi Fractale en Imagerie medicale : outils, validation et application », Thèse de Doctorat, Université de Lille1, 2009
- [7] JACOPO Grazzini, « Analyses multi échelle et multi fractale d'images météorologiques : Application à la détection de zones précipitantes » Human-Computer Interaction [cs.HC]. Université de Marne la Vallée, 2003.
- [8] S.ZEHANI (1), M.KHIDER (2), A.TALEB AHMED (3), M.MIMI (4), « ANALYSE FRACTALE DE TEXTURE : APPLICATION A L'IMAGE IRM ET CT-SCAN DE L'OS TRABECULAIRE », Université de Biskra, Algerie, 2011
- [9] MENZE et al. « THE MULTIMODAL BRAIN TUMOR IMAGE SEGMENTATION BENCHMARK (BRATS) », OCTOBER 2015
- [10] Nicholas Ayache. L'analyse automatique des images medicales : Etat de l'art et perspectives. RR-3364, INRIA. 1998.
- [11] RONAN Fable, « F4B202, Analyse de texture », Département Signal & Communications, Telecom Bretagne
- [12] ALICE Porebski, « Classification d'images couleur texturées », Rapport de stage, UNIVERSITE DES SCIENCES & TECHNOLOGIES DE LILLE, 2006
- [13] LEHAMEL Malha, « Segmentation d'images texturées à partir des attributs fractals », thèse de Magister, Université de Tizi Ouzou,
- [14] LAURENT Risser, « ANALYSE MULTI-ECHELLE DES TEXTURES », 2003

- [15] Gérard Morvan, Valérie Vuillemin, Henri Guerini, Philippe Mathieu, Marc Wybier, Frédéric Zeitoun, Philippe Bossard, Patrick Stérin, « Imagerie Médicale LEONARD DE VINCI »
- [16] Society For Neuroscience & Société des Neurosciences, 2013
- [17]<https://www.espci.fr/fr/actualites/2014/une-nouvelle-technique-d-imagerie-cerebrale-pour>.
- [18] BATY Xavier, « Classification tissulaire par analyse texturale en IRM cardiaque du petit animal », mémoire de Stage, LISA FRE 2656 Université d'Angers, France
- [19] AMROUN Fazia, « Extraction de la composante texturée d'une image », Thèse de Magister, Université de Tizi ouzou, 2013
- [20] Marie Chavent, « Bagging et Forêts aléatoires, Chapitre 4 », Université de Bordeaux, 2015-2016
- [21] Michèle Gouiffès, « cours d'Analyse de Texture », Université Paris sud 11, France
- [22] Alceu Ferraz Costa, Gabriel Humpire-Mamani, Agma Juci Machado Traina, « An Efficient Algorithm for Fractal Analysis of Textures, SFTA », University of São Paulo, USP, São Carlos, Brazil
- [23] Robin Genuer, « Forêts aléatoires : aspects théoriques, sélection de variables et applications », Thèse de Doctorat, Université Paris Sud 11, 2010