

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



Université 20 Août 1955-Skikda  
Faculté des Sciences  
Département d'Informatique



Mémoire de Fin d'Etude  
En Vue de l'Obtention du Diplôme  
De Master en Informatique  
Option « Systèmes Informatiques SI »

**Thème :**



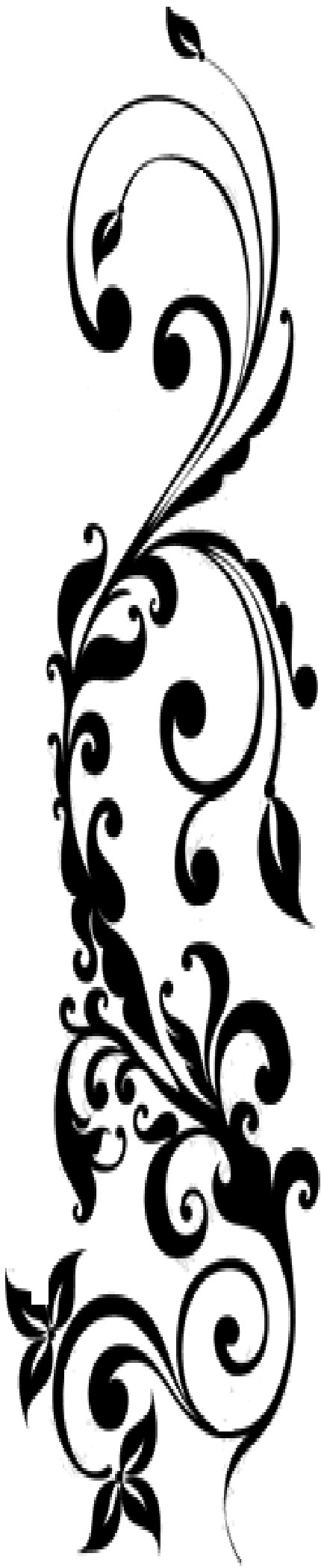
**Techniques d'Apprentissage  
Automatique Pour la  
Segmentation D'Images Médicales  
(IRM)**

**Réalisé Par :** Bourekkouk Samira et Denden Sihem

**Devant le jury de soutenance**

- Mme Belaid Hassina, Université de Skikda, Présidente
- Mme Megroun Hanan, Université de Skikda, Examinatrice
- M. Mazouzi Smaine, Université de Skikda, Encadreur

Session 1 , 2023



# *Remerciements*

# *Remerciement*

*Nous remercions Allah, le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce Modeste travail.*

*Nous adressons notre reconnaissance, notre gratitude À Notre Enseignant et Encadreur et Vice Recteur le professeur*

***Mr /Maazouzi Smain***

*Nous tenons ,tout particulièrement et très sincèrement, à lui remercier de nous avoir proposé le sujet et de nous avoir encadrés , son suivi , ses encouragements et ses orientations et son aide précieuse, Qu'il nous soit permis d'exprimer nos plus vifs remerciements de nous avoir fait bénéficier de ses compétences, ses qualités humaines et de sa disponibilité non seulement pour la réalisation de ce mémoire mais aussi durant tout le parcours de notre formation.*

*Ainsi qu'à la réussite de cette formidable Cycle de Master.*

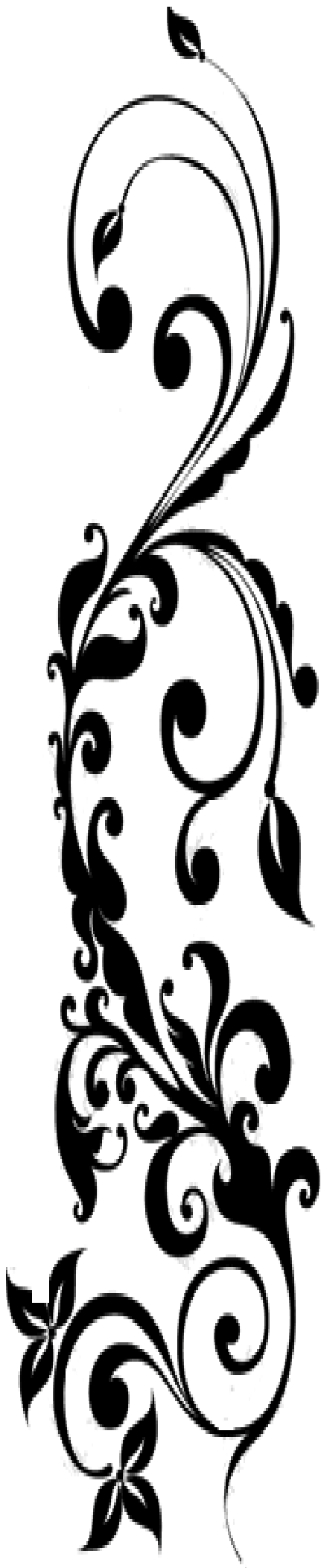
*Nos remerciements s'adressent également au nombre de Jury pour l'honneur qu'ils nous font d'avoir bien voulu étudier ce Travail et de le juger.*

*Nous n'oublions pas d'adresser un grand merci au corps professoral et administratif du département , pour la richesse et la qualité de leur enseignement et pour les grands efforts pour assurer à leurs étudiants une formation actualisée.*

*Nous souhaitant adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire.*

*Merci à Tous .*





# *Dédicace*

# Dédicace

À mes très chers parents  
Pour tout leur amour et toute leur tendresse  
Je prie **ALLAH** pour les protéger,  
À mon mari : **Tarek** qui m'a aidé, encouragé et a été un  
soutien pour moi,

À mes cher fils : **Mohamed Anes** et ma petite **Lyna**,  
À mes chères sœurs et Frères chacun par son nom,  
À la famille de mon mari, qui m'ont donné la force et la  
détermination pour réussir.

À mes respectés enseignants, qui m'ont apporté le savoir,  
l'orientation et la sagesse tout au long de ce voyage éducatif,

À tous mes camarades d'étude et de travail qui ont  
partagé avec moi des jours remplis de défis et de  
réussites.

À mes collègues de travail : « service des personnels,  
technique et agents de services » et « sous direction des  
personnels et formation » au sein du rectorat de  
l'université de Skikda ,

Avec lesquelles j'ai partagé d'agréables moments.  
À tous ceux qui sont chères, proches de mon cœur, et à  
tous ceux qui m'aiment et qui aurait voulu

**partager ma joie**



*Samira*

# Dédicace

*À mes très chers parents*

*Pour tout leur amour et toute leur tendresse*

*Je prie ALLAH pour les protéger*

*À mon mari : **Toufik** qui m'a aidé, encouragé et été un soutien pour moi, "*

*À mon cher fils : **Abdelalim***

*À toutes mes familles et à la famille de mon mari, qui m'ont donné la force et la détermination pour réussir.*

*À mes respectés enseignants, qui m'ont apporté le savoir, l'orientation et la sagesse tout au long de ce voyage éducatif,*

*À tous mes camarades d'étude et de travail qui ont partagé avec moi des jours remplis de défis et de réussites.*

*À mes amies: Hanen , Nawel et Sabrina.*

*Avec lesquelles j'ai partagé d'agréables moments.*

*À tous ceux qui sont chères, proches de mon cœur, et à tous ceux qui m'aiment et qui aurait voulu*

*partager ma joie*



*Sihem*

# Abstract

Image segmentation, and in particular in medical images, remains problematic due to noise and various distortions. In this master's thesis, we propose a new method of segmenting brain MRIs by pixel classification, adopting a supervised approach using the k-means algorithm. The proposed method is based on the classification of pixels in local regions, in which the problem of non-uniformity of intensities (INU: Intensity Non Uniformity) can be neglected. However, the partial results must be merged using a majority voting technique.

# Résumé

*La segmentation d'images et en particulier pour les images médicales, demeure problématique à cause du bruit et de différentes déformations. Dans ce travail de master, nous proposons une nouvelle méthode de segmentation d'IRM cérébrales par classification des pixels, en adoptant une approche supervisée en utilisant l'algorithme des moyennes mobiles (k-means). La méthode proposée est basée sur la classification des pixels dans des régions locales, dans lesquelles le problème la non-uniformité des intensités (INU : Intensity Non Uniformity) pourra être négligé. Cependant les résultats partiels doivent être fusionnés en utilisant une technique de vote majoritaire.*

# ملخص

لا يزال تجزئة الصور و خاصة الصور الطبية ، يمثل مشكلة بسبب التشويش و التشوهات المختلفة . في أطروحة الماستر هاته نقتراح طريقة جديدة لتجزئة الصور بالرنين المغناطيسي للدماغ حسب تصنيف البيكسل، باتباع نهج خاضع للإشراف باستخدام خوارزمية المتوسطات المتحركة (k-mean) ، نعتمد الطريقة المقترحة على تصنيف وحدات بيكسال في المناطق المحلية، حيث يمكن اهمال مشكلة عدم انتظام الشدة (UNI : Intensity Non Uniformity) و مع ذلك يجب دمج النتائج الجزئية باستخدام تقنية التصويت بالأغلبية.

# Sommaire:

Désignation	Pages
Abstract	
Résumé	
ملخص	
Sommaire	
Table de figure	
Introduction générale	01
<b>Chapitre 1: Segmentation d'Images médicale</b>	<b>02</b>
Introduction	03
<b>Partie 1 : Imagerie médicales</b>	<b>03</b>
1-Définition de l'anatomie	03
1-1-Le cerveau	03
1-1-1-La matière grise	04
1-1-2-La matière blanche	04
1-1-3-Le liquide céphalo-rachidien	05
1-2-Observation du cerveau	05
2- Imagerie médicale	06
2-1-Techniques de l'imagerie médicale	07
3-Imagerie par résonance magnétique (IRM)	07
3-1-Principe général de L'IRM	08
3-2-Principales composantes d'IRM	09
3-3-Acquisition des images IRM	11
<b>Partie 2 : Segmentation d'image</b>	<b>14</b>
1-Définition de la segmentation	14
2-Les différentes Méthodes de segmentation	14
..	
2-1-Approches Contours	16
2-1-1-Les méthodes dérivatives	16
2-1-2-Méthodes déformables	17
2-2-Approche région	18
2-2-1-Croissance de région (région growing)	18
2-2-2-Segmentation par fusion de régions (Merge)	19
2-2-3-Segmentation par division de régions (Split)	19

2-2-4-Segmentation par division-fusion (Split and Merge)	19
2-3-Segmentation par classification	20
2-3-1-Méthodes non supervisées	20
2-3-2-Méthodes supervisées	22
2-3-3-Méthode d'ensemble de classifieurs	22
2-3-3-1-Production d'un ensemble de classifieurs	23
2-3-3-2-Intégration d'un ensemble de classifieurs	23
2-3-3-3-Évaluation d'un modèle en apprentissage automatique	23
3-État de l'art sur quelques techniques de segmentation d'images IRM	25
4-Quelques difficultés de la segmentation des images médicales	28
5-Conclusion	30
<b>Chapitre 2 : Apprentissage Automatique</b>	<b>31</b>
1-Introduction	32
2-L'apprentissage automatique	32
2-1-Définition	32
2-2 -Modélisation	32
3-Domains de l'apprentissage automatique	33
3-1- La fouille de données et ses méthodes	34
3-2- L'intelligence Artificielle	36
4-Types de l'apprentissage automatique	36
4-1-Apprentissage supervisé	37
4-2-Apprentissage non supervisé	38
4-3-Apprentissage semi-supervisé	38
4-4-Apprentissage par renforcement	39
5-Les principaux algorithmes d'apprentissage automatique	40
5-1- Non paramétrique	40
5-1-1- Les k plus proches voisins	40
5-1-2- Les arbres de décisions	41
5-2-Paramétrique	42
5-2-1- Classification naïve bayésienne	42
5-2-2- Les machines à vecteurs de support SVM	43
6-Classification non-supervisée par k-mens	44
7-Les méthodes d'agrégation Boosting et Bagging	44
7-1- Boosting	44
7-2-Bagging	45
8-Conclusion	46

<b>Chapitre 3 : Moyennes mobiles locales pour la segmentation d'IRM cérébrales</b>	<b>47</b>
1- Introduction	48
2- Algorithme k-means	48
2-1- Définition	48
2-2- Principe de la méthode des K-Means	49
2-3- Principe Algorithmique	51
2-4- Organigramme de l'algorithme de k-means	52
2-5-Convergence et initialisation des K-Means	52
2-6- Choix du nombre K de classes (clusters)	53
2-7-Cas d'utilisation k-means	53
3-Principe de la méthode proposée	53
4- Étapes de la méthode	55
4-1-Filtrage de bruit	55
4-2-Sélection de régions aléatoires	55
4-3- Application locale de k-means	56
4-4-Fusion des résultats	56
4-5-Finalisation	57
5-Conclusion	58
<b>Chapitre 4: Implémentation et Test</b>	<b>59</b>
1- Introduction	60
2- Le langage Python	60
2-1-A propos du langage Python	60
2-2-Quelles sont les principales raisons qui poussent à apprendre Python	61
2-3 -Bibliothèques en Python	62
2-4-Bibliothèques utilisés	62
3-Environnement de développement	63
3-1-Matériel utilisé	63
3-2-Outils de développement	63
3.2.1- Google Colab	63
3-2-2-Les offres de Google Colab	64
4-Création de projet et l'explication de l'interface	64
4-1-Extraits du code	68
4-2-Chargement, filtrage et affichage IRM	68
4-3-Chargement données d'apprentissage	69
4-4-Initialisation des données d'apprentissage et sélection des sous images	69

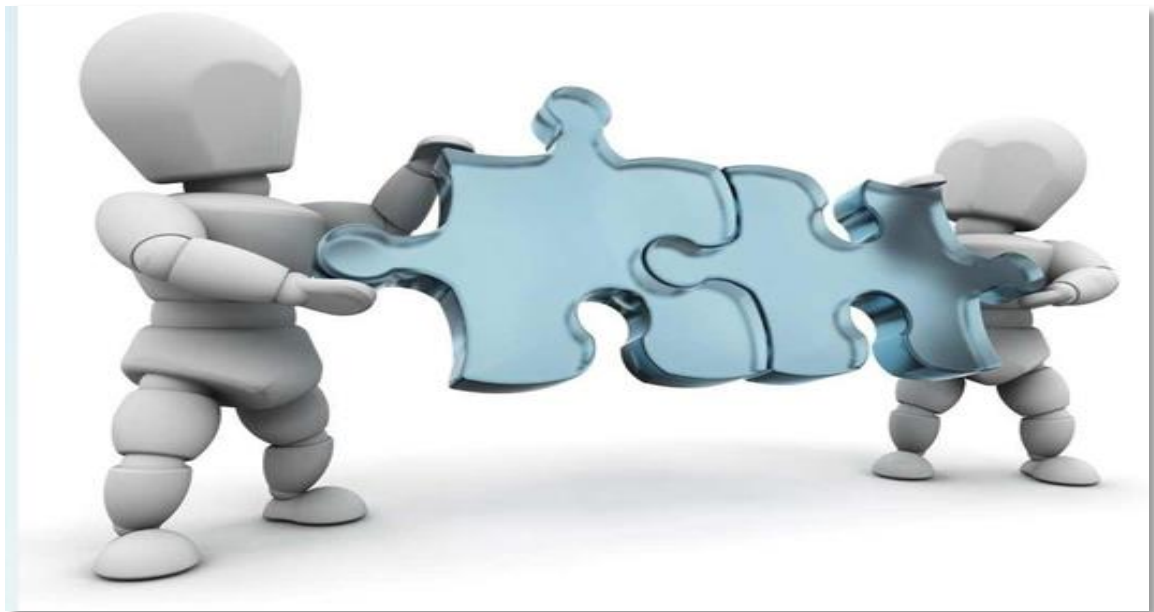
4-5-K-meand local et copie des résultats	<b>70</b>
4-6-Affichage du résultat global	<b>70</b>
4-7-Exemple de matrice de confusion	<b>70</b>
4-8-Exemple de Métriques d'évaluation	<b>71</b>
4-9-Exemples de résultats de k-means locaux	<b>71</b>
4-10-Résultat global	<b>73</b>
5-Conclusion	<b>73</b>
Conclusion Générale	<b>75</b>
Bibliographie	<b>76</b>

## Table des figures :

<b>N° Figure</b>	<b>Désignation de figures</b>	<b>Pages</b>
<b>Figure chapitre 1</b>		
<b>Figure I.01</b>	Image général du cerveau humain	<b>04</b>
<b>Figure I.02</b>	L'image illustre l'encéphale, composée de principales parties et d'une coupe d'un cerveau mettant en évidence les trois matières principales.	<b>05</b>
<b>Figure I.03</b>	Les trois axes de coupe pour la visualisation du cerveau et l'IRM cérébrale	<b>06</b>
<b>Figure I.04</b>	L'appareil IRM	<b>08</b>
<b>Figure I.05</b>	Image composite les trois axes X, Y et Z de coupe pour la visualisation du cerveau en IRM	<b>10</b>
<b>Figure I.06</b>	Principales composantes d'IRM	<b>11</b>
<b>Figure I.07</b>	Acquisition des images IRM du cerveau : coup axiale pondération T1-T2 le liquide du système ventriculaire (flèche) est hyperintense	<b>12</b>
<b>Figure I.08</b>	Synthèse T1-T2 : TR (Temps de Répétition) et TE (Temps d'Écho)	<b>13</b>
<b>Figure I.09</b>	Organigramme Général des approches de segmentation	<b>15</b>
<b>Figure I.10</b>	Segmentation par approche contour	<b>16</b>
<b>Figure I.11</b>	Contour et ses dérivées.	<b>16</b>
<b>Figure I.12</b>	Exemple de l'approche contour déformable	<b>17</b>
<b>Figure I.13</b>	Segmentation d'images par accroissement de région	<b>18</b>
<b>Figure I.14</b>	Décompositions successives des blocs.	<b>19</b>
<b>Figure I.15</b>	Exemple d'une matrice de confusion.	<b>24</b>
<b>Figure I.16</b>	Résultats de segmentation	<b>26</b>
<b>Figure I.17</b>	Segmentation probabiliste en 3 classes de 2 types de séquences d'acquisition.	<b>27</b>
<b>Figure I.18</b>	Exemple d'image pondérée en T affectée par l'artefact d'inhomogénéité de champ magnétique radiofréquence. (a) Image originelle. (b) Carte d'estimation de l'inhomogénéité de champ magnétique B. (c) Image corrigée.	<b>28</b>
<b>Figure I.19</b>	(a) Schéma représentatif en 2D d'une structure cérébrale imagée à travers une grille de pixels. Les pixels rouges et jaunes sont sujets aux effets de volume partiel.	<b>29</b>
<b>Figure chapitre 2</b>		
<b>Figure II.01</b>	Schéma de modélisation d'une machine d'apprentissage	<b>33</b>
<b>Figure II.02</b>	Schéma des différentes techniques issues de l'IA et FD pour la construction de modèles de données	<b>33</b>
<b>Figure II.03</b>	Processus d'extraction de connaissances à partir des données	<b>34</b>
<b>Figure II.04</b>	Les différents types d'apprentissage automatique	<b>36</b>

<b>Figure II.05</b>	Schéma d'un modèle supervisé	<b>37</b>
<b>Figure II.06</b>	Schéma d'un modèle non supervisé	<b>38</b>
<b>Figure II.07</b>	Schéma d'un modèle semi-supervisé	<b>39</b>
<b>Figure II.08</b>	Schéma d'un modèle par renforcement	<b>39</b>
<b>Figure II.09</b>	Fonctionnement de k plus proche voisins	<b>41</b>
<b>Figure II.10</b>	Architecture parallèle de SVM	<b>43</b>
<b>Figure II.11</b>	Fonctionnement du Boosting	<b>45</b>
<b>Figure II.12</b>	Fonctionnement du Bagging	<b>46</b>
<b>Figure chapitre 3</b>		
<b>Figure III.1</b>	Classification à base de K-means	<b>49</b>
<b>Figure III.2</b>	Exemple d'initialisation du K-Means avec trois clusters	<b>49</b>
<b>Figure III.3</b>	Mise à jour des centres après une itération de K-means	<b>50</b>
<b>Figure III.4</b>	Exemples de clusters finaux obtenus avec K-means	<b>50</b>
<b>Figure III.5</b>	Organigramme de l'Algorithme K-Means	<b>52</b>
<b>Figure III.6</b>	Non uniformité d'intensité dans les IRM	<b>54</b>
<b>Figure III.7</b>	Déroulement de la méthode proposée	<b>54</b>
<b>Figure chapitre 4</b>		
<b>Figure IV.1</b>	Création un projet dans Google Colab	<b>65</b>
<b>Figure IV.2</b>	Partie gauche de l'interface	<b>65</b>
<b>Figure IV.3</b>	Partie droite de l'interface	<b>67</b>
<b>Figure IV.4</b>	Image brute à segmenter	<b>69</b>
<b>Figure IV.5</b>	03 Exemple des k-means locaux	<b>72</b>
<b>Figure IV.6</b>	Image global segmentée par k-means	<b>73</b>

# Introduction Général



# Introduction Générale

L'imagerie médicale est utilisée à large échelle par les professionnels de santé pour établir le diagnostic médical pour la quasi-totalité des pathologies. Les modalités en imagerie médicale sont nombreuses allant de la simple échographie et en arrivant aux techniques nucléaires telles que la scintigraphie et l'imagerie par émission de positrons.

Par ailleurs, la segmentation d'images médicales est une étape primordiale pour tout post-traitement d'analyse d'images. Elle consiste en général partitionner l'image traitée en ses différentes parties, structurelles, fonctionnelles ou simplement radiométriques. Cependant les images médicales sont souvent entachées de bruit et de différentes déformations ce qui rend difficile leur segmentation.

En imagerie par résonance magnétique (IRM), le problème de la non-uniformité de l'intensité (INU : Intensity Non Uniformity) qui consiste à une variation de l'intensité, du au dispositif physique d'acquisition est un problème difficile qui s'ajoute au problème de bruit. Pour remédier à ce problème, nous proposons dans ce mémoire une nouvelle méthode de segmentation des tissus dans les IRM cérébraux. Le principe de la méthode consiste à classifier les pixels de l'image en opérant dans une région locale dans laquelle le problème de l'INU est réduit au maximum pourra être négligé. Ensuite, fusionner les résultats obtenus dans les différentes régions locales.

Pour la classification, nous avons considéré l'approche non supervisée, et nous avons utilisé l'algorithme des moyennes mobiles (k-means).

Ce mémoire est organisé comme suit :

**Au chapitre 1**, nous survolons le domaine de l'imagerie médicale, en présentant les différentes modalités d'images médicales, leurs caractéristiques, et les technologies de leur acquisition.

**Le chapitre 2** est consacré aux méthodes d'apprentissage automatique, où nous allons concentrer sur les méthodes supervisées, dont nous allons utiliser la méthode des moyennes mobiles (k-means).

**Le chapitre 3** présente en détail notre méthode, en donnant son principe et nous présentons ses différentes étapes. Nous montrons comment les régions locales sont générées et comment les classes fiables des pixels sont calculées.

**Le chapitre 4** est consacré à l'implémentation et les tests de la méthode proposée dans lequel nous commençons par présenter l'environnement et le langage de programmation à savoir python Ensuite, nous montrons quelques interfaces de l'application, et quelques résultats expérimentaux avec discussion.

**Une conclusion générale** résume notre travail, et souligne ses perspectives.

# CHAPITRE 01 :

## Segmentation D'Images Médicales

# CHAPITRE 1 : Segmentation d'Images médicale

## Introduction :

L'imagerie médicale est un élément essentiel à la recherche clinique, l'étude des maladies et la mise au point de nouveaux traitements. Il existe de nombreuses techniques d'imagerie ciblant à explorer les différents organes et fonctions du corps humain.

La segmentation est le cœur d'un système d'analyse automatique d'images , ainsi est un ensemble d'algorithmes utilisés pour distinguer des objets d'intérêt, c'est à-dire trouvé des caractéristiques locales adéquates permettant de les séparer des autres objets ou du fond. La segmentation est une étape fondamentale dans le processus du traitement d'image en raison de ses nombreux domaines d'application [1].

Dans ce chapitre, La première partie, nous présentons l'anatomie cérébrale et l'imagerie médicales IRM. La deuxième partie, est consacrée pour la segmentation d'images qui se considère parmi les opérations les plus importantes dans un système de traitement d'images. Ensuite nous allons citer quelques publications concernant la segmentation d'images médicales IRM, et à la fin de ce chapitre on discutera sur les différents problèmes de segmentation.

## **Partie 1 : Imagerie Médicales**

### **1-Définition de l'anatomie:**

L'anatomie est la science qui a pour objet l'étude de la structure et de la morphologie des êtres vivants et en particulier de l'homme , c'est l'étude de la structure des organes par leur dissection. Elle est fondée sur l'observation et la description minutieuse des structures des corps vivants. Elle recherche la connaissance des structures qui composent les organismes vivants en précisant leurs situations, leurs formes, leurs rapports, leurs fonctions, leurs particularités, etc. Longtemps construite à partir de dissections post-mortem, elle utilise aujourd'hui d'autres techniques d'investigation (radiographie, tomographie, échographie, endoscopie, imagerie par résonance magnétique, prélèvements tissulaires, etc.) [2]

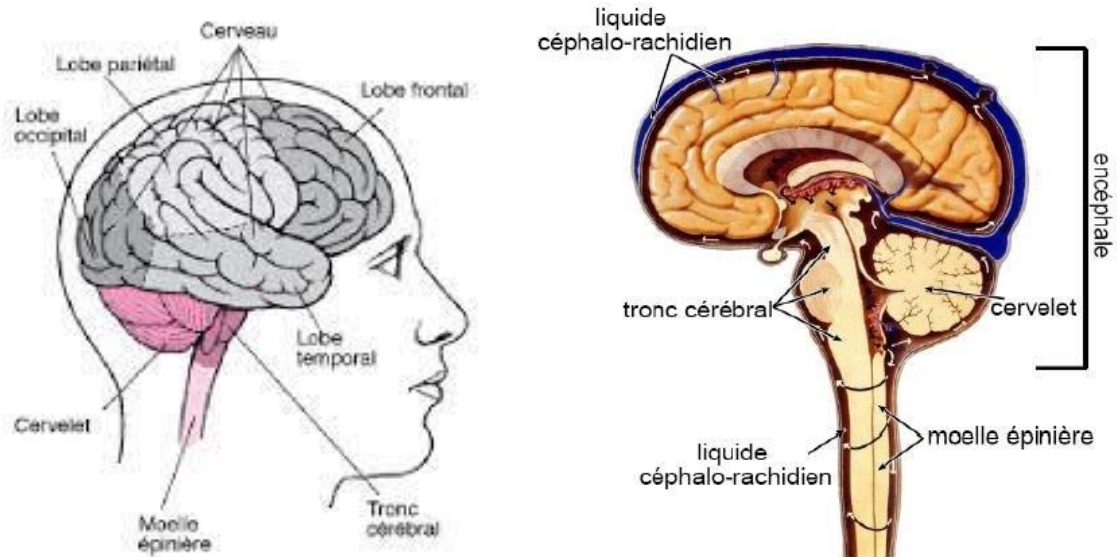
### **1-1 - Le cerveau :**

Le système nerveux central est l'organe qui contrôle l'ensemble des comportements de l'individu, qu'il s'agisse de comportements conscients ou inconscients. Il est constitué d'un réseau complexe de communications dont les influx règlent le fonctionnement des organes. Anatomiquement, le système nerveux central est composé de trois parties : la moelle épinière, le tronc cérébral et le cerveau. Il est la partie la plus volumineuse du système nerveux central. Il est constitué de deux hémisphères, séparés par la scissure inter hémisphérique, et reliés par différentes

# CHAPITRE 1 : Segmentation d'Images médicale

structures telles que le corps calleux, le thalamus et l'hypothalamus. Le cerveau est composé de substance (ou matière) grise externe appelée cortex, de substance blanche sous-jacente et de structures profondes appelées noyaux gris centraux. Le cerveau baigne dans un liquide appelé liquide céphalorachidien [3].

Le cerveau est composé de trois tissus principaux : le liquide céphalo-rachidien, la matière grise et la matière blanche.



**Figure I.01 : Image général du cerveau humain**

## 1-1-1 - La matière grise :

On peut distinguer plusieurs structures composées de matière grise. D'une part, on trouve le cortex, situé à la périphérie du cerveau et d'autre part les noyaux gris centraux. La matière grise est essentiellement composée de neurones et constitue le siège de l'activité cérébrale.

**a. Le cortex :** Le cortex recouvre la totalité du cerveau. Sa surface est importante car il suit toutes les convolutions externes du cerveau, appelées sillons. Son épaisseur est d'environ 2 à 3 millimètres.

**b. Les noyaux gris centraux :** Ces noyaux sont également composés de matière grise. Les principaux noyaux sont le noyau caudé, qui longe le ventricule latéral, le noyau lenticulaire composé du putamen et du pallidum et le noyau thalamique qui est le plus gros des noyaux centraux[3].

## 1-1-2 - La matière blanche :

La substance blanche est un tissu de couleur blanchâtre qui fait partie du système nerveux central. Située dans le cerveau et la moelle épinière, elle est constituée d'axones (les prolongements des neurones

# CHAPITRE 1 : Segmentation d'Images médicale

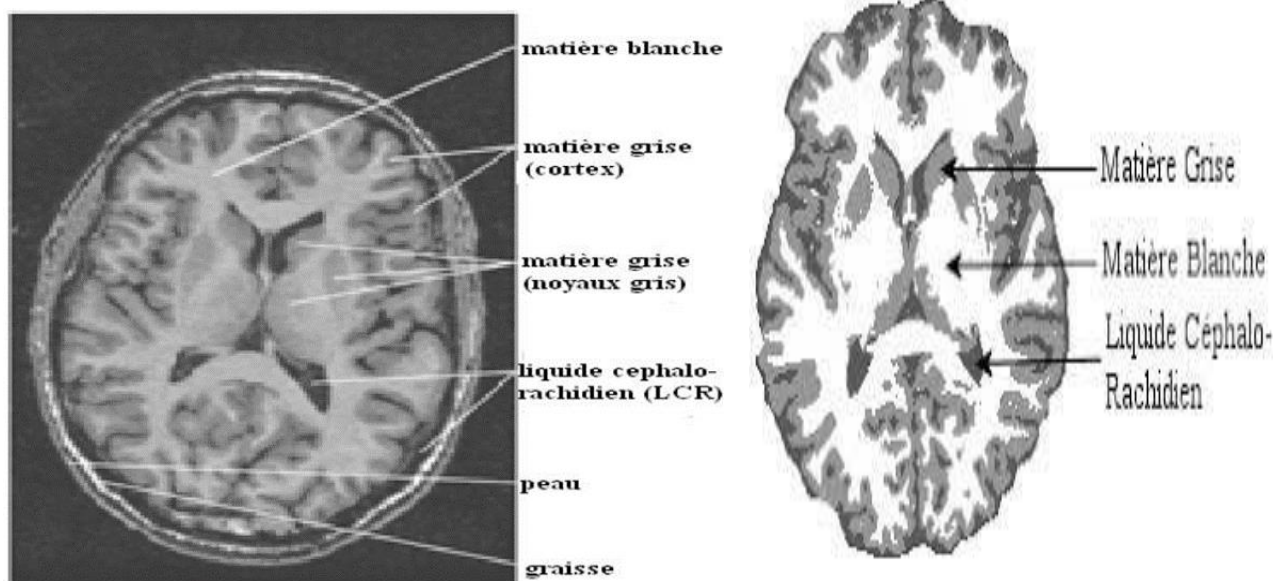
qui transportent les signaux électriques jusqu'aux zones synaptiques) et de la névroglie (tissu qui soutient les cellules nerveuses). Le rôle de la substance blanche est d'assurer la bonne circulation des informations dans le système nerveux [4].

## 1-1-3 - Le liquide céphalo-rachidien (LCR) :

Le liquide céphalo-rachidien ou cérébro-spinal (en anglais cerebro-spinalfluid) entoure le cerveau et remplit les ventricules cérébraux. Il permet de protéger le cerveau de s chocs en jouant un rôle d'amortisseur, et remplit des fonctions importantes dans les échanges entre le sang et les tissus nerveux [3].

## 1-2-Observation du cerveau :

L'observation par des coupes en trois dimensions (3D) du cerveau peut être effectuée selon plusieurs angles de vue. Ainsi, on distingue trois axes anatomiques pour réaliser les coupes illustrées dans la Figure 2



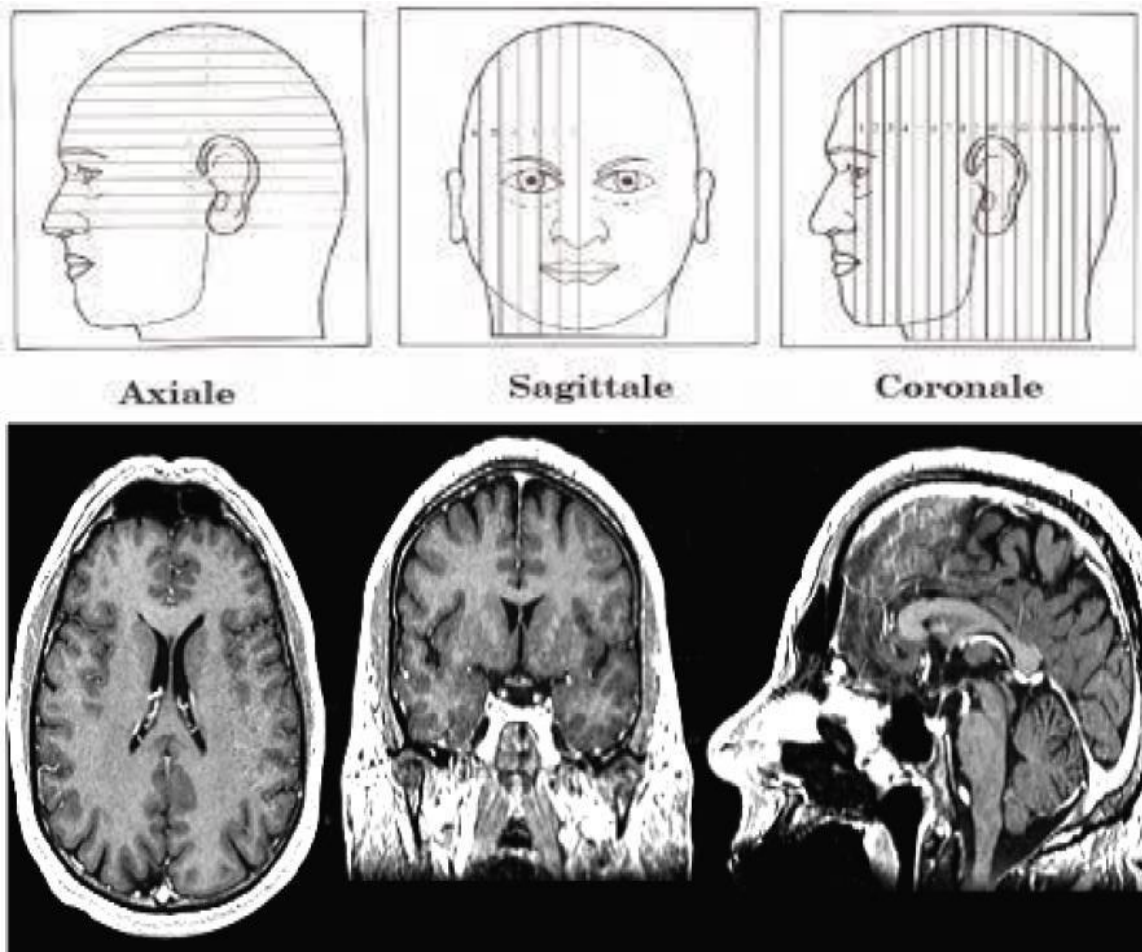
**Figure I.02 : L'image illustre l'encéphale, composée de principales parties et d'une coupe d'un cerveau mettant en évidence les trois matières principales.**

**1-Coupes axiales :** Ces coupes correspondent quasiment à un plan horizontal. En imagerie de résonance magnétique, elles correspondent à un plan perpendiculaire à l'axe du champ magnétique principal.

**2-Coupes sagittales :** Ces coupes sont prises dans des plans parallèles au plan inter hémisphérique. Il s'agit de vues latérales du cerveau.

**3-Coupes coronales:** Ce sont des coupes perpendiculaires aux coupes axiales et sagittales. Elles correspondent à des vues de face du cerveau [3].

# CHAPITRE 1 : Segmentation d'Images médicale



**Figure I.03:** Les trois axes de coupe pour la visualisation du cerveau et l'IRM cérébrale

## 2-Imagerie médicale :

L'imagerie médicale est certainement l'un des domaines de la médecine qui a le plus progressé ces vingt dernières années. Elle regroupe l'ensemble des techniques et les moyens d'acquisition et de restauration images du corps humain à partir de différents phénomènes physiques qui consistent à mettre en image différentes régions ou différents organes de l'organisme à partir de différents types de rayonnements qui visualise une partie du corps humain ou d'un organe sans avoir à opérer le patient, cela en créant une image visuelle compréhensible d'une information à caractère médical dans le but d'établir un diagnostic et de faire un suivi approprié du traitement [4,5].

- ❖ **L'imagerie structurelle** : permet d'avoir des informations sur l'anatomie des organes (taille, volume, localisation et forme éventuelle d'une lésion).
- ❖ **L'imagerie fonctionnelle** : permet d'avoir des informations sur le fonctionnement des organes (physiologie, métabolisme, ...) [6].

# CHAPITRE 1 : Segmentation d'Images médicale

L'importance que revêt l'imagerie médicale tient d'abord au fait qu'une image est un concentré d'information bien plus efficace qu'un texte ou qu'une explication verbale. L'interprétation des images médicales est un des domaines de recherche les plus encourageants, étant donné qu'il offre des facilités pour le diagnostic et les décisions thérapeutiques d'un grand nombre de maladies .

## 2-1-Techniques de l'imagerie médicale

Les différentes modalités d'imagerie médicale sont nombreuses et permettent d'obtenir des informations différentes selon le procédé physique utilisé pour observer les tissus du corps humain. Certains procédés apportant une information anatomique tandis que d'autres détectent les fonctionnalités des organes en offrent une carte d'activités. Durant de nombreuses années ces images étaient obtenues par la radiographie conventionnelle, avec l'avancé du progrès, l'imagerie médicale connaît un renouveau technologique considérable et le développement de techniques nouvelles telles que :

- Les Rayons X;
- L'échographie;
- La Tomographie par Émission de Positons (PET);
- Le scanner (TDM);
- La résonance magnétique (IRM).

Une des techniques d'imagerie les plus couramment utilisées, à laquelle nous nous sommes intéressés dans le cadre de ce travail, est l'Imagerie par Résonance Magnétique (IRM), qui est devenue un outil indispensable pour tout examen clinique. Dans ce qui suit nous présentons le principe général d'IRM.

## 3-Imagerie par résonance magnétique (IRM)

L'imagerie par Résonance Magnétique, ou est une technique de diagnostic médical puissante qui fournit des images tridimensionnelles et en coupe de grande précision anatomique. L'IRM est une technique radiologique récente, non invasive et sans effets secondaires connus, basée sur le phénomène physique de résonance magnétique nucléaire. Il s'agit simplement d'observer la résonance magnétique nucléaire (RMN) des protons de l'eau contenus dans l'organisme, l'organisme, c'est à dire la réponse des noyaux soumis à un champ magnétique extérieur et à une excitation électromagnétique [7].

# CHAPITRE 1 : Segmentation d'Images médicale

## 3-1-Principe général de L'IRM

L'imagerie par résonance magnétique nucléaire (IRM) est une technique d'imager médicale d'apparition récente, elle est basée sur le phénomène physique de résonance magnétienucléaire. Il s'agit simplement d'observer la résonance magnétique nucléaire (RMN) des protons de l'eau contenus dans l'organisme. Le principe consiste à mesurer l'aimantation des tissus biologiques et à reconstruire une image à partir de ces aimantations. Cette dernière est induite par l'hydrogène, qui se trouve en abondance dans le corps humain. Le patient Placés dans un puissant champ magnétique, tous les atomes d'hydrogène s'orientent dans la même direction : ils sont alors excités par des ondes radio durant une très courte période. À l'arrêt de cette stimulation, les atomes restituent l'énergie accumulée en produisant un signal qui est enregistré et traité sous forme d'image par un système informatique [8].

Après arrêt de la stimulation, les atomes d'hydrogène restituent cette énergie qui se dissipe dans différents plans de l'espace sous l'action du champ magnétique de l'aiment. L'énergie est alors captée par d'autres antennes (antennes réceptrices) puis analysée par un ordinateur. Suivant la composante en eau des tissus analysés, leurs pathologies éventuelles, les images seront différentes et l'ordinateur réalise des images en noir et blanc d'une très grande sensibilité qui seront très précieuses pour le diagnostic, notamment en matière de pathologie tumorale ou infectieuse. Il est ainsi possible de réaliser des coupes de n'importe quelle partie du corps dans tous les plans de l'espace [10].



**Figure I.04: L'appareil IRM**

# CHAPITRE 1 : Segmentation d'Images médicale

## 3-2-Principales composantes d'IRM

Un système d'imagerie par résonance magnétique est constitué principalement des éléments suivants :

**a. L'aimant :** est au cœur du fonctionnement de l'appareil IRM. Son rôle est de produire le champ magnétique principal appelé  $\vec{B}$  qui est constant et permanent. L'unité de mesure de la puissance du champ magnétique de l'IRM est le Tesla [8].

**b- Les bobines de gradient de champ magnétique:** Ce sont trois bobines métalliques enfermées dans un cylindre en fibres de verre et placées autour du tunnel de l'aimant. On les nomme respectivement : bobine X, bobines Y et bobine Z.

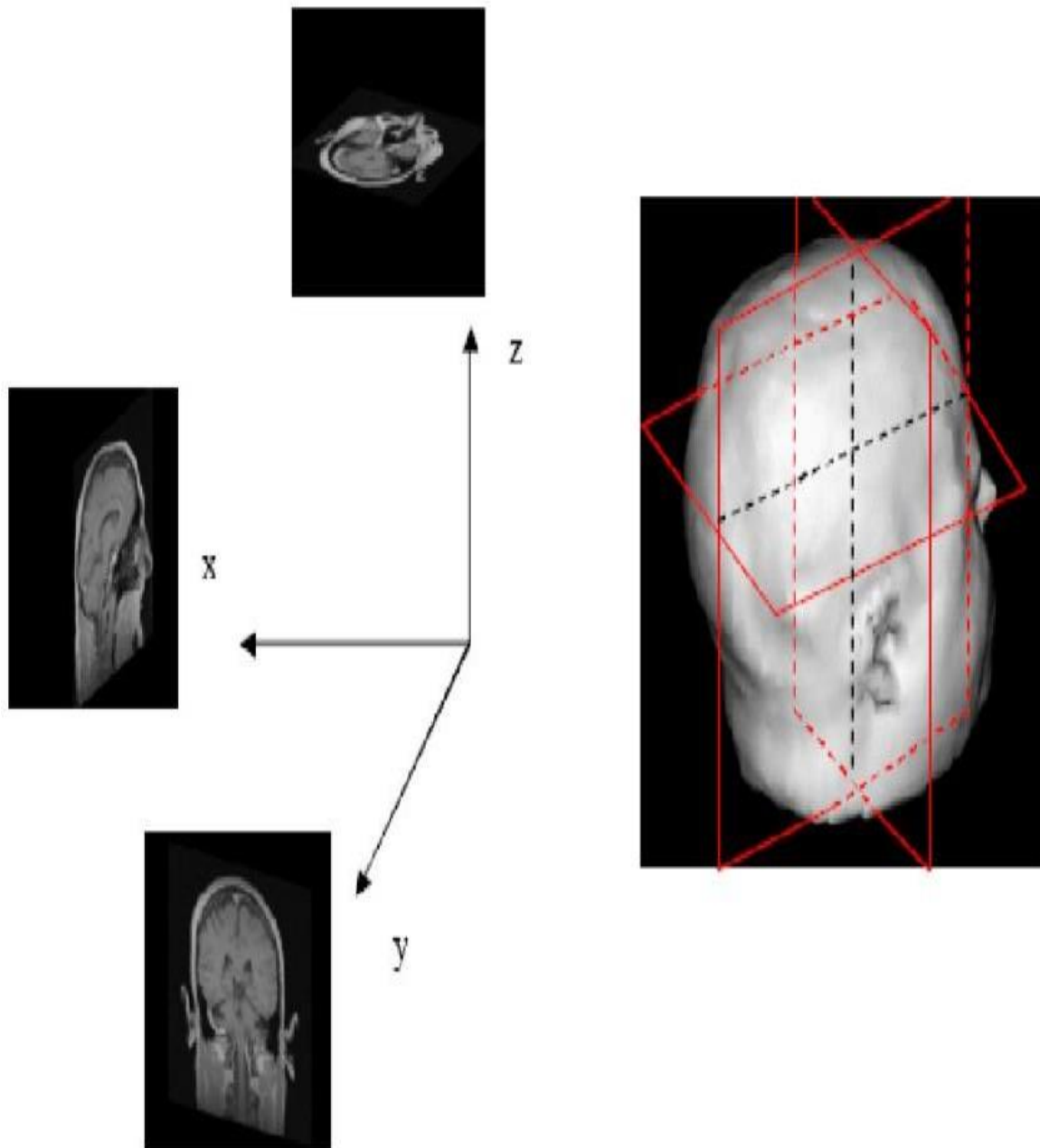
Le passage d'un courant électrique dans ces bobines crée des variations d'intensité du champ magnétique dans le tunnel, de façon linéaire, dans le temps et dans l'espace. En fonction de sa géométrie, chaque bobine fait varier le champ magnétique selon un axespécifique :

- la bobine X selon l'axe droite-gauche
- la bobine Y selon l'axe avant-arrière
- la bobine Z selon l'axe haut-bas.

Elles permettent notamment de sélectionner une épaisseur et un plan de coupe (transversal, frontal, sagittal ou oblique) et d'effectuer la localisation spatiale des signaux dans ce plan [8].

La visualisation des différentes incidences de coupe en imagerie médicale l'axe X correspondant l'incidence sagittale ; l'axe Y à l'incidence Coronal ; et l'axe Z à l'incidence Axiale lors de l'utilisation de données issues de l'IRM (Figure 04).

## CHAPITRE 1 : Segmentation d'Images médicale



**Figure I.05:** Image composite les trois axes X, Y et Z de coupe pour la visualisation du cerveau en IRM

**c. Les antennes :** Elles sont capables de produire et/ou capter un signal de radiofréquence (R.F.). Elles sont accordées pour correspondre à la fréquence de résonance de précision des protons qui se trouvent dans le champ magnétique, les antennes sont très variables [8].

**d. Le calculateur:** un processeur vectoriel assure la gestion de l'ensemble de l'acquisition et du traitement. Il commute les gradients, contrôle les séquences d'impulsions, recueille les signaux, effectue la transformée de Fourier(FFT) et stocke sur un support magnétique l'image sur une matrice de pixels [9].

# CHAPITRE 1 : Segmentation d'Images médicale

e. **Les blindages:** en IRM, on parle de blindages pour certains dispositifs destinés au confinement des champs magnétiques produits par la machine et à l'isolement de celui-ci des champs magnétiques extérieurs qui viendraient perturber l'acquisition. Il permet des acquisitions synchronisées sur un autre signal périodique (comme l'ECG, le rythme respiratoire) [8].

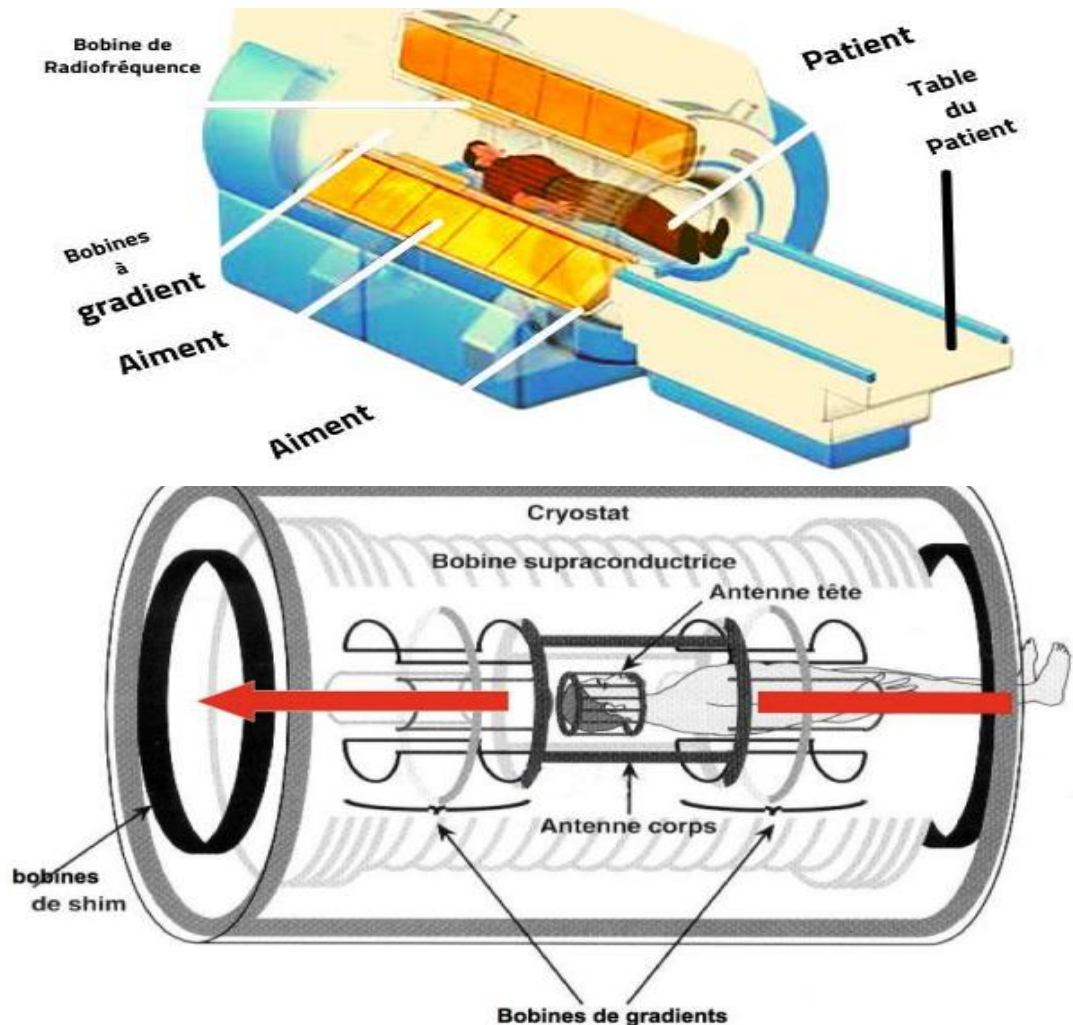


Figure I.06: Principales composantes d'IRM

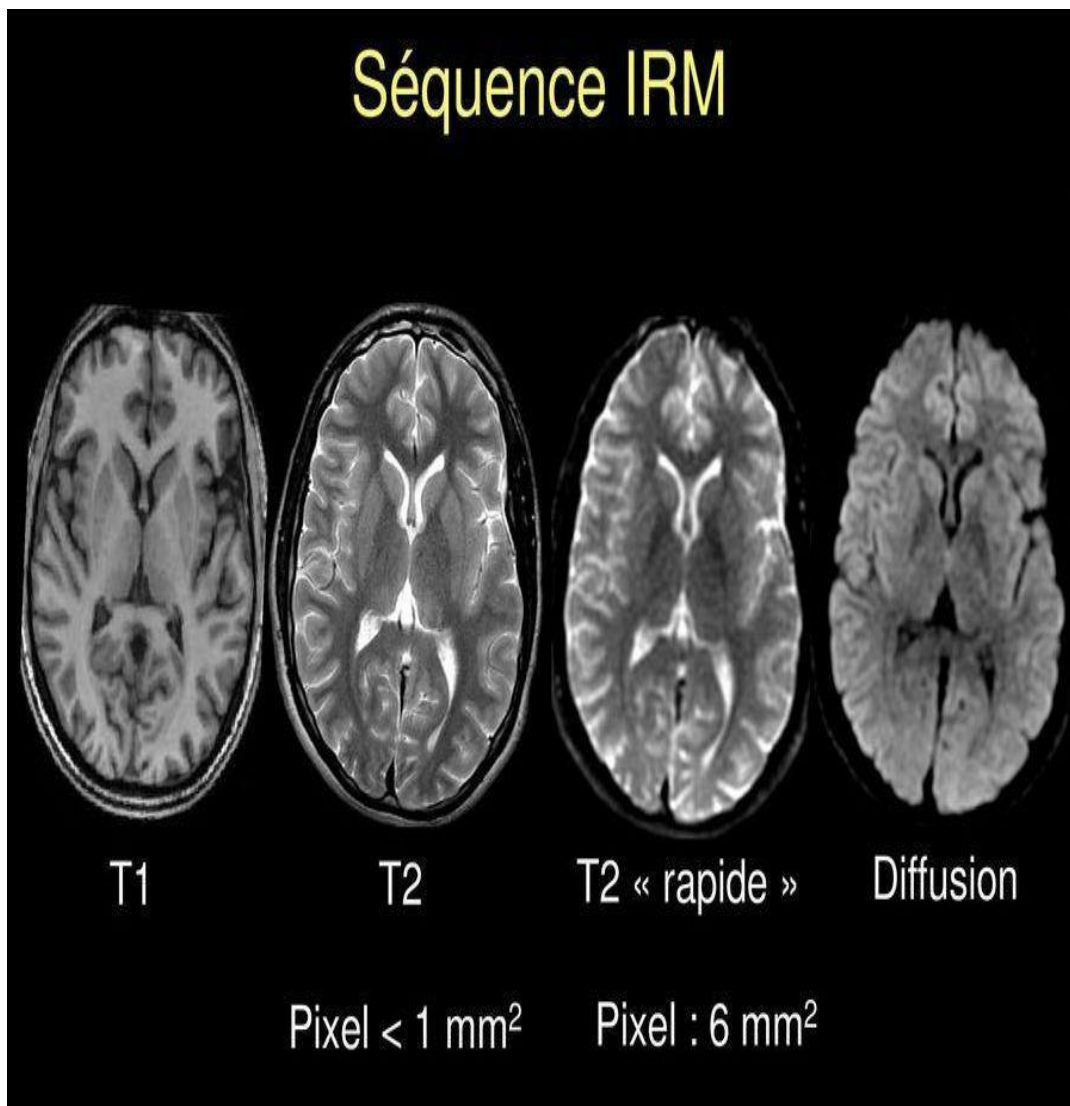
## 3-3-Acquisition des images IRM

L'IRM permet d'obtenir des images numériques en trois dimensions d'une précision inférieure au millimètre, du cortex, de la substance blanche, du liquide céphalo-rachidien et des noyaux gris centraux [8].

Deux paramètres réglables cliniquement entrent en jeu lors de l'acquisition : le temps de répétition des séquences d'impulsion (TR), et le temps d'écho (TE). Une acquisition effectuée avec des temps TR et TE courts est dite en T1 (Figure 07 et 08).

## CHAPITRE 1 : Segmentation d'Images médicale

Lorsque ces deux temps sont longs, elle est dite en T2 (Figure 07 et 08). Enfin lorsque (TR) est long et (TE) est court, elle est dite en rho ou densité de proton. Toute l'acquisition peut être caractérisée par rapport à ces trois types particuliers. Ainsi, on parle d'image pondérée en T1 si l'acquisition se rapproche plus de type d'acquisition en T1 que des deux autres types : de façon similaire, on parle d'acquisition pondérée en T2 ou en  $\rho$  (Figure 08). De plus, on peut renforcer le signal de certaines structures sur les séquences pondérées en T1 grâce à l'utilisation d'un agent de contraste tel que le gadolinium [10].



**Figure I.07 : Acquisition des images IRM du cerveau : coup axiale pondération T1-T2 le liquide du système ventriculaire (flèche) est hyperintens**

Au niveau cérébral, l'IRM permet également d'étudier les artères du cou et du cerveau comme nous l'avons vu précédemment les différentes coupes du cerveau, Coupe axiale, Coupe coronale et Coupe sagittale sur la figure 03.

# CHAPITRE 1 : Segmentation d'Images médicale

## Synthèse T1-T2 : TR (Temps de Répétition) et TE (Temps d'Écho)

	pT1	pT2
TR	Court	Long
TE	Court	Long E1-E2

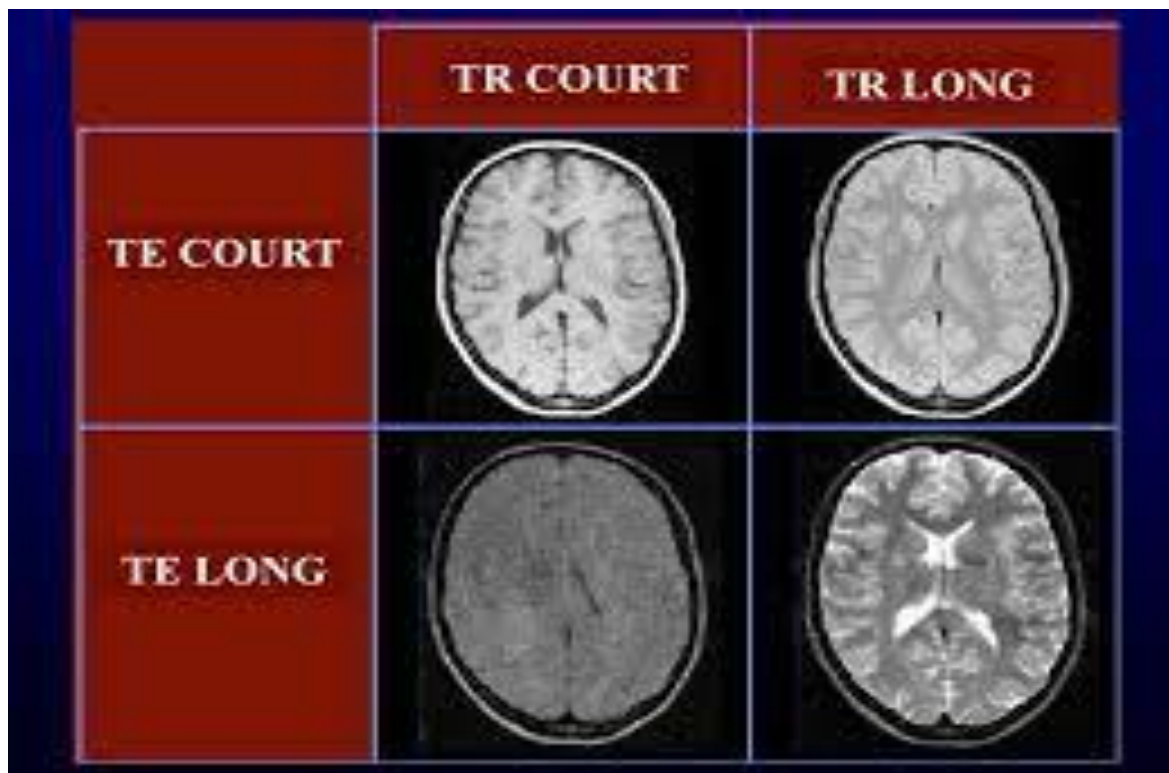


Figure I.08 : Synthèse T1-T2 : TR (Temps de Répétition) et TE (Temps d'Écho)

# CHAPITRE 1 : Segmentation d'Images médicale

## Partie 2: Segmentation d'image

### 1-Définition de la segmentation

La segmentation est une des étapes critiques de l'analyse d'images qui conditionne la qualité des mesures effectuées ultérieurement. C'est généralement une première étape d'un traitement plus complexe comme la reconnaissance de formes. Elle permet de cerner les formes des objets sur lesquels doit porter l'analyse, de délimiter des régions, l'intérêt et de les extraire du fond. Une bonne méthode de segmentation sera celle qui permettra d'arriver à une bonne interprétation. Elle devra donc avoir simplifié l'image sans pour autant en avoir trop réduit le contenu[11].

Segmenter, ou délimiter les structures cérébrales est une étape fondamentale pour l'analyse d'images cérébrales. Elle permet de faire une séparation des différents tissus cérébraux (matière grise, matière blanche, liquide céphalorachidien, etc..) ainsi que d'éventuelles pathologies cérébrales. Une bonne segmentation permet d'aider le médecin à prendre une décision finale, avant son geste chirurgical.

De nombreuses études des tissus du cerveau dans les images d'IRM ont été effectuées, telles que des techniques basées contours, région, la logique floue, des méthodes statistiques, et bien d'autres, nous présenterons les principales méthodes de segmentation utilisées dans la suite du chapitre.

### 2-Les différentes Méthodes de segmentations

# CHAPITRE 1 : Segmentation d'Images médicale



Figure I.09 : Organigramme Général des Méthodes de segmentation  
p. 15

# CHAPITRE 1 : Segmentation d'Images médicale

## 2-1-Approche Contours

L'approche contour consiste à identifier les changements entre les régions. En général, un élément de contours est un point de l'image appartenant à la frontière de deux ou plusieurs objets ayant des niveaux de gris différents. Les variations d'intensité de lumière et de couleurs sont très bien perçues par le système visuel humain [12].

Il existe plusieurs méthodes utilisant l'approche contours, dans ce qui suit, nous allons présenter les plus importantes.



Figure I.10 : Segmentation par approche contour

### 2-1-1-Les méthodes dérivatives

Les méthodes dérivatives sont les plus utilisées pour détecter des transitions d'intensité par différenciation numérique (Première et deuxième dérivé). À chaque position, un opérateur est appliqué afin de détecter les transitions significatives au niveau de l'attribut de discontinuité choisi. Le résultat est une image binaire constituée de points de contours et de points non - contours [12].

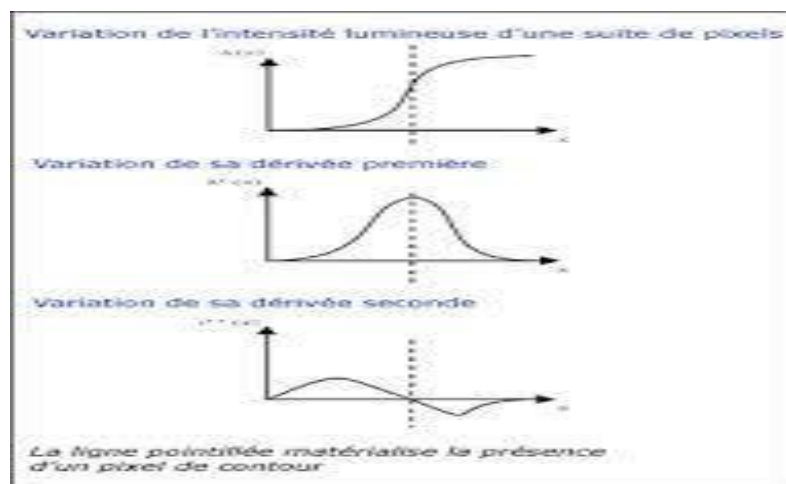


Figure I.11 : Contour et ses dérivées.

# CHAPITRE 1 : Segmentation d'Images médicale

## 2-1-2-Méthodes déformables :

Les modèles déformables sont aussi connus sous les noms de « *snakes* » ou « contours actifs ». L'intérêt principal des contours actifs est de détecter des objets dans une image en utilisant les techniques d'évolution de courbes. L'idée est de partir d'une courbe initiale, généralement un carré ou un cercle, et de la déformer jusqu'à obtenir le contour de l'objet. En effet, celui-ci présente quelques inconvénients tels que la sensibilité à l'initialisation, au bruit, et le réglage difficile de ses différents paramètres.

Les contours actifs tirent leur origine des modèles élastiques, mais la communauté s'accorde à les attribuer à l'équipe Kass, Witkin et Terzopoulos [13], qui introduisirent les Snakes ou courbes minimisantes. Les snakes tiennent leur nom de leur aptitude à se déformer comme des serpents.

❖ **Les principales limites** des méthodes de détection de contour sont les suivantes [14] :

- Les techniques de détection de contour dépendent de l'information contenue dans le voisinage local de l'image. Il n'y a pas d'information globale.
- Après l'extraction des points de contours, ces derniers sont reliés afin de déterminer les frontières. Le processus de fermeture des contours peut parfois conduire à des discontinuités et des lacunes dans l'image.
- Il est souvent difficile d'identifier et de classer les contours parasites.

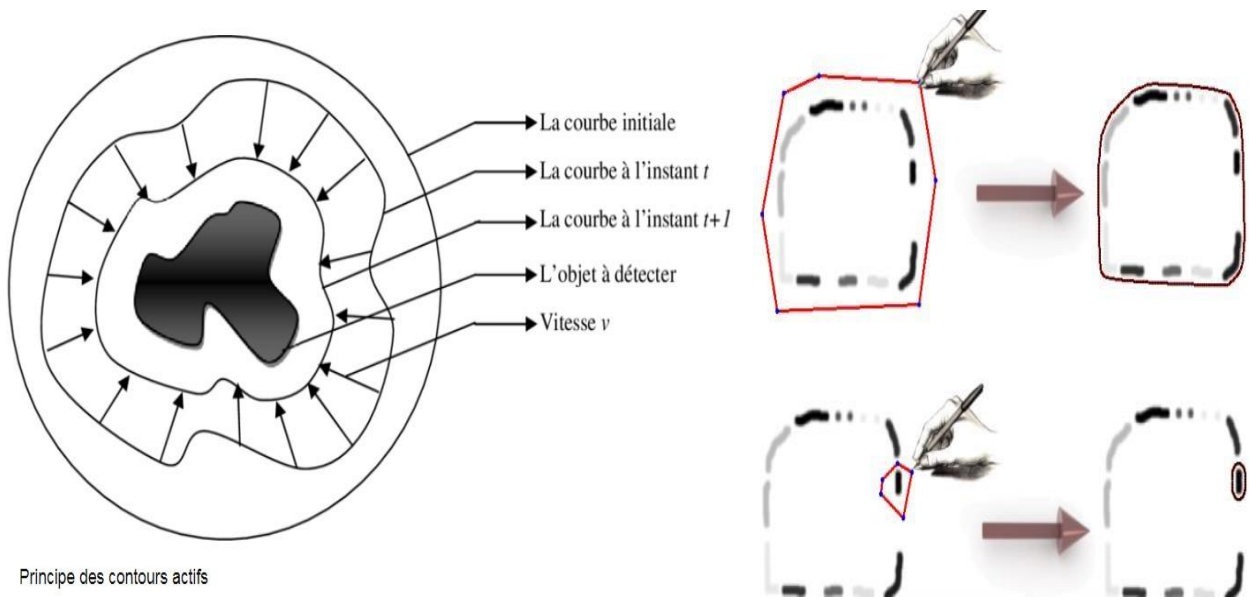


Figure I.12: Exemple de l'approche contour déformable

# CHAPITRE 1 : Segmentation d'Images médicale

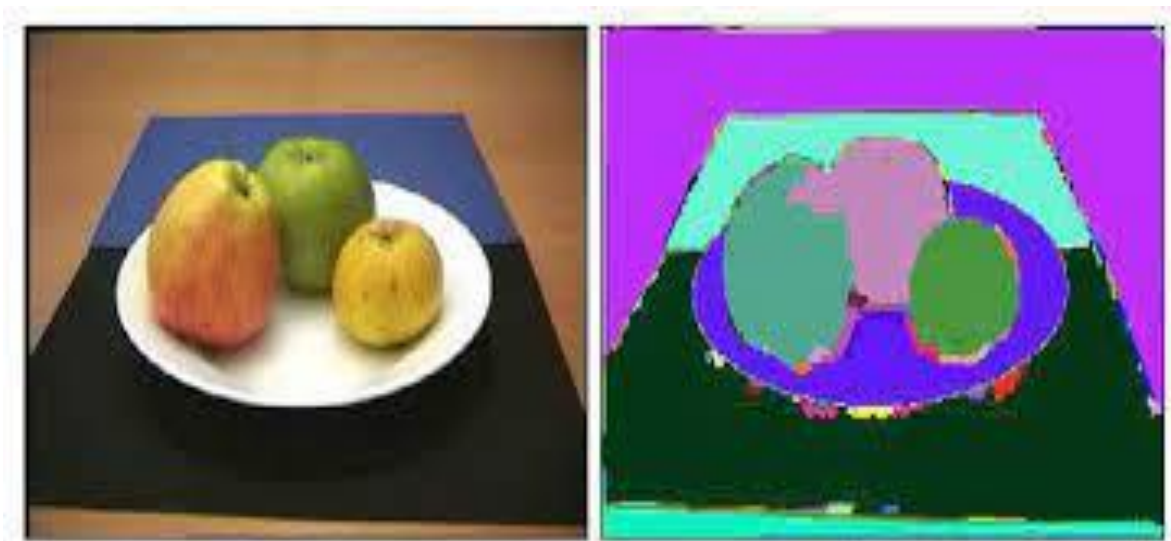
## 2-2-Approche Région :

La segmentation par régions consiste à décomposer l'image en régions homogènes, Ces régions sont composées d'ensemble des pixels connexes possédant des propriétés au sens d'un prédicat d'homogénéité Donné. Ce critère peut être, par exemple, le niveau de gris, la couleur, la texture...etc. Un processus de groupement est répété jusqu'à ce que tous les pixels dans l'image soient inclus dans des régions. Cette approche vise, donc, à segmenter l'image en se basant sur des propriétés intrinsèques des régions [12]. Il existe plusieurs méthodes telles que la segmentation par croissance de région, division de région, par fusion de région et par division fusion que nous présentons ci-dessous.

### 2-2-1- Croissance de région (région growing) :

Cette technique consiste à faire progressivement accroître les régions autour de leur point de départ. L'initialisation de cette méthode consiste à considérer chaque pixel comme une région. On va essayer de les regrouper entre elles avec un double critère de similarité des niveaux de gris et d'adjacence. Le critère de similarité peut par exemple être : la variance des niveaux de gris de la région R est inférieure à un seuil.

Le principe de l'agrégation de pixel est le suivant : on choisit un germe (Le point de départ est le choix d'un ensemble de pixels appelés « germes ») et on fait croître ce germe tant que des pixels de son voisinage vérifient le test d'homogénéité. Lorsqu'il n'y a plus de pixels candidats dans le voisinage, on choisit un nouveau germe et on itère le processus [15].



**Figure I.13: Segmentation d'images par accroissement de région**

# CHAPITRE 1 : Segmentation d'Images médicale

## 2-2-2-Segmentation par fusion de régions (Merge) :

Les techniques de réunion (region merging) sont des méthodes ascendantes où tous les pixels sont visités. Pour chaque voisinage de pixel, un prédicat P est testé. S'il est vérifié les pixels correspondants sont regroupés dans une région [15].

Les inconvénients de cette méthode se situent à deux niveaux :

- Cette méthode dépend du critère de fusion qui peut influencer sur le résultat final de la segmentation.
- Elle peut introduire l'effet de sous- segmentation.

## 2-2-3-Segmentation par division de régions (Split) :

La division consiste à partitionner l'image en régions homogènes selon un critère donné. Le principe de cette technique est de considérer l'image elle-même comme région initiale, qui par la suite est divisée en régions. Le processus de division est réitéré sur chaque nouvelle région (issue de la division) jusqu'à l'obtention de classes homogènes [13].

Cette méthode présente un inconvénient majeur qui est la sur-segmentation. Toute fois, ce problème peut être résolu en utilisant la méthode de division-fusion que nous présentons dans ce qui suit.

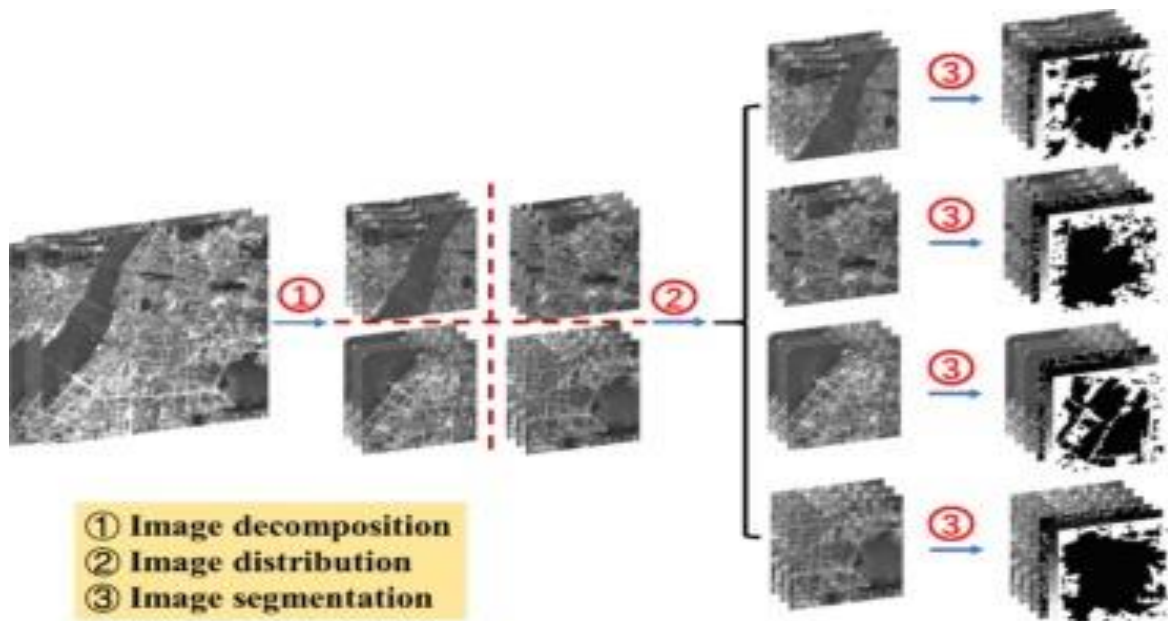


Figure I.14 : Décompositions successives des blocs.

## 2-2-4-Segmentation par division-fusion (Split and Merge)

Ces méthodes combinent les deux méthodes décrites précédemment, la division de l'image en de petites régions homogènes, puis la fusion des régions connexes et similaires au sens d'un prédicat de regroupement. On part du principe que chaque pixel représente à lui seul une région. Deux régions seront fusionnées si elles répondent aux critères de similarité des niveaux de gris et

# CHAPITRE 1 : Segmentation d'Images médicale

d'adjacence de régions .On s'arrête quand le critère de fusion n'est plus vérifié [12].

Les inconvénients de la segmentation par région se situent à trois niveaux :

- Les régions obtenues ne correspondent pas, dans tous les cas, aux objets représentés dans l'image.
- Les limites des régions obtenues sont habituellement imprécises et ne coïncident pas exactement aux limites des objets de l'image.
- La difficulté d'identifier les critères pour agréger les pixels ou pour fusionner et diviser les régions.

## 2-3-Segmentation par classification :

Les méthodes de classification permettent de regrouper des objets en groupes ou classes d'objets plus homogènes. Les objets regroupés ont des caractéristiques communes, ils sont similaires mais se distinguent clairement des objets des autres classes .Généralement les résultats des méthodes de classification sont plus ou moins différents.

La classification peut être supervisée ou non supervisée.

### 2-3-1-Méthodes non Supervisées :

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune base d'apprentissage et par là même aucune tâche préalable d'étiquetage manuel. La seule intervention de l'expert se situe à la fin du processus pour identifier les images. Elles sont utilisées pour effectuer une classification en aveugle et, ainsi pour réaliser une segmentation sans connaissances a priori sur l'image.

Il est par conséquent nécessaire de prendre quelques précautions. La mesure de proximité est calculée sur tous les attributs, il n'y a donc pas de réduction de l'espace de représentation. Comme il n'y a pas d'exemples d'apprentissage, si certains attributs utilisés sont non représentatifs des classes recherchées, le résultat pourra être médiocre.

On peut citer les algorithmes de classification non supervisé comme: k-means, Fuzzy c- means.

#### Algorithmes K-means :

L'algorithme K-means (k-moyenne ou centre mobiles) est simple et largement utilisé. C'est l'un des algorithmes les plus connus en classification non supervisé. Il a été mis en point par « MacQueen » à la fin des années soixante (1967). Il est basé sur la méthode des centroïde. (centres de gravité). Le principe de cette méthode est de choisir un nombre K de classes de l'ensemble d'arrivée.

Le nombre K est choisi selon une connaissance préalable : combien de région sont souhaitées. On

# CHAPITRE 1 : Segmentation d'Images médicale

affecte alors à chaque pixel l'étiquette de la classe le plus proche (en prenant par exemple la distance euclidienne comme critère d'affectation) puis les centres de gravité sont recalculés. Le processus est ainsi réitéré jusqu'à atteindre un état de stabilité ou aucune amélioration n'est possible.

La méthode des k-moyennes est très attrayante pour obtenir rapidement une première classification. De plus, c'est une méthode simple et rapide. Cependant, pour le cas particulier des images médicales, la supposition a priori par ces méthodes que les limites entre les classes sont très bien définies peut poser problème. Pour pallier ces limites, une autre méthode qui repose sur la logique floue est née c'est l'algorithme des c-moyennes floues (*Fuzzy C-means*) [13].

## ➤ **Les avantages de Algorithmes K-means :**

- Très populaire du fait qu'il est très facile à comprendre et à mettre en œuvre.
- la méthode résolve une tâche non supervisée, donc elle ne nécessite aucune information sur les données.
- Rapidité et faibles exigences en taille mémoire.
- La méthode est applicable à tout type de données (même textuelles), en choisissant une bonne notion de distance.

## ➤ **Inconvénients de Algorithmes K-means :**

- Le nombre de classes est un paramètre de l'algorithme. Un bon choix du nombre K est nécessaire, car un mauvais choix de K produit de mauvais résultats.
- les points isolés sont mal gérés (doivent-ils appartenir obligatoirement à un cluster ?)
- L'algorithme du K-Means ne trouve pas nécessairement la configuration la plus optimale correspondant à la fonction objective minimale.
- Les résultats de l'algorithme du K-Means sont sensibles à l'initialisation aléatoire des centres.

## **Fuzzy C-means :**

Fuzzy C-Means (FCM) est un algorithme de classification non-supervisée floue. Issu de l'algorithme des C-moyennes (C-means), il introduit la notion d'ensemble flou dans la définition des classes : chaque point dans l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leur centre de gravité. Comme les autres algorithmes de classification non supervisée, il utilise un critère de minimisation des distances intra-classe et de maximisation des distances interclasse, mais en donnant un certain degré d'appartenance à chaque classe pour chaque pixel. Cet algorithme nécessite la connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective. Ainsi, il permet

# CHAPITRE 1 : Segmentation d'Images médicale

d'obtenir une partition floue de l'image en donnant à chaque pixel un degré d'appartenance (compris entre 0 et 1) à une classe donnée. Le cluster auquel est associé un pixel est celui dont le degré d'appartenance sera le plus élevé. Les principales étapes de l'algorithme Fuzzy C-means sont :

- La fixation arbitraire d'une matrice d'appartenance.
- Le calcul des centroïdes des classes.
- Le réajustement de la matrice d'appartenance suivant la position des centroïdes.
- Calcul du critère de minimisation et retour à l'étape 2 s'il y a non convergence de critère [8].

La principale faiblesse de cet algorithme se trouve lors de l'initialisation des centres, aussi ces méthodes sont sensibles au bruit qui malheureusement toujours présent dans les images IRM.

## 2-3-2-Méthodes supervisées

Ce sont des méthodes dans lesquelles les classes sont connues a priori avant d'effectuer l'opération d'identification des éléments de l'image. Elles demandent une phase d'apprentissage sur l'échantillon représentatif dans le but d'apprendre les caractéristiques de chaque classe et une autre phase pour décider l'appartenance d'un pixel à telle ou telle classe. Parmi ces méthodes on peut citer : la segmentation Bayésienne, la segmentation par les champs de Markov , réseaux de neurones ...etc. Pour notre cas, puisqu'il s'agit de segmenter des images médicales IRM cérébrales, la classification supervisée de ces images nécessite donc la création d'une base d'apprentissage pour chaque classe et pour chaque patient ce qui est en elle-même une tâche très fastidieuse pour l'expert [16]. C'est pour cette raison que nous intéressons aux méthodes non supervisées.

## 2-3-3-Méthode d'ensemble de classifieurs :

Les méthodes de classification que nous avons présentée précédemment, consistent à construire un classifieur unique. Toutefois, de nombreuses études empiriques confirment que la performance des méthodes d'ensemble qui consistent à combiner plusieurs classifieurs simples et locaux est généralement meilleure que celle d'un seul classifieur élémentaire de base. On n'observe souvent que la performance en généralisation (sur des nouveaux exemples) d'un système reposant sur un ensemble de classifieurs continue à augmenter lorsqu'on augmente le nombre de classifieurs, et cela, même lorsque la performance en apprentissage semble avoir atteint le maximum [17].

### L'intérêt de la combinaison de classifieurs :

- Distribuer les caractéristiques sur des classifieurs adaptés;
- Exploiter la complémentarité entre classifieurs;
- Prendre en compte les performances de chacun des classifieurs;

# CHAPITRE 1 : Segmentation d'Images médicale

- Réduire l'importance des choix initiaux;
- Diviser pour mieux régner.

En général, la conception d'un système de classification à base d'ensemble de classifieurs comporte généralement trois étapes: la définition et la mise au point des classifieurs et le choix de la méthode d'intégration ou de fusion de ces classifieurs, et l'évaluation.

## 2-3-3-1-Production d'un ensemble de classifieurs :

La première étape est donc de produire un ensemble de classifieurs élémentaires performants. Dans cette étape on cherchera aussi à obtenir une bonne diversité entre les classifieurs afin que les décisions fournies par les classifieurs soient peu corrélées qui doivent être évaluées sur différentes parties de l'ensemble d'apprentissage. La construction d'un ensemble de classifieurs est généralement réalisée par les techniques suivantes :

- ✓ Soit en utilisant des algorithmes d'apprentissage différents;
- ✓ Ou bien en utilisant le même algorithme mais avec des paramètres ou initialisations différents;
- ✓ Ou d'une autre part en utilisant des sous-ensembles d'apprentissage différents avec le même algorithme. Parmi toutes les approches, les plus connues, le bagging, le boosting ' [17].

## 2-3-3-2-Intégration d'un ensemble de classifieurs :

L'intégration des sorties des classifieurs élémentaires ainsi générés est un autre aspect important intervenant dans les performances du classifieur final.

Le vote est la technique d'intégration la plus couramment utilisée où la prédiction finale est obtenue en comptabilisant les sorties (votes) des classificateurs élémentaires. La forme la plus simple est le vote majoritaire, dans lequel chaque classifieur apporte un seul vote pour une classe particulière et la prédiction finale se fait en faveur de la classe qui a reçu une majorité de votes.

Une version réside dans le vote majoritaire pondéré, dans lequel l'influence de chaque classifieur sur la prédiction finale est pondérée par sa performance estimée sur un ensemble de validation. La prédiction finale est alors de choisir la classe ayant la plus grande valeur de la somme pondéré des votes. Ces techniques apparaissent comme simples et efficaces [17].

## 2-3-3-3-Évaluation d'un modèle en apprentissage automatique :

Un modèle statistique de classification va nous servir à étiqueter (classer) des nouveaux individus non étiquetés. Pour créer ce modèle, nous utilisons des individus déjà étiquetés. Afin de tester la qualité de ce modèle, nous réservons aussi une une partie de nos individus étiquetés pour des « tests » et une partie « apprentissage », en se basant sur la validation croisée («cross-validation »).

# CHAPITRE 1 : Segmentation d'Images médicale

La validation croisée est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. En fait, il y a au moins trois techniques de validation croisée : « testset validation » ou « holdoutmethod », « k-fold cross-validation » et « leave-one-out cross-validation ».

## *k-fold cross-validation :*

C'est la méthode choisie pour ce travail. On divise l'échantillon original en  $k$  échantillons, puis on sélectionne un des  $k$  échantillons comme ensemble de validation et les  $(k-1)$  autres échantillons constitueront l'ensemble d'apprentissage. L'opération se répète ainsi  $k$  fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation.

Ainsi, on va classer les individus « tests » en faisant comme si on ne connaissait pas leur étiquette en les passant dans le classifieur. Enfin, on va comparer ces nouvelles étiquettes avec les vraies étiquettes pour connaître le taux de bonne classification et donc la qualité du modèle [17].

En croisant, les étiquettes réelles des individus avec leur nouvelle étiquette prédite par le modèle, on va pouvoir mettre en place une matrice de confusion.

## *La matrice de confusion*

Pour expliquer ce qu'est une matrice de confusion, voici un exemple :

Imaginons que nous souhaitons créer un modèle statistique permettant de prédire le temps qu'il fait dehors en fonction de relevés météorologiques. Les individus statistiques sont les villes, les variables explicatives sont des relevés de température, pression atmosphérique, luminosité et les classes à prédire sont « pluie », « beau temps » et « neige ».

Admettons qu'on ait fait tout le travail en amont permettant de créer le modèle statistique avec les données d'apprentissage et on a prédit les classes des données de tests.

On va donc croiser ces données et obtenir une matrice de ce genre dans la figure 14 :

	Pluie	Beau Temps	Neige
Pluie	31	1	9
Beau Temps	6	23	8
Neige	5	6	32

**Figure I.15: Exemple d'une matrice de confusion.**

# CHAPITRE 1 : Segmentation d'Images médicale

Comme on peut l'observer, une matrice de confusion a en colonne et en ligne les mêmes intitulés et elle n'est pas pour autant une matrice symétrique.

En ligne, on lit les labels des individus et en colonne les labels prédits par le modèle. Ainsi, on peut conclure par exemple que :

- ✓ Dans 31 situations où il pleuvait, le modèle a bien prédit qu'il pleuvait
- ✓ Le modèle a prédit 1 fois qu'il faisait beau temps alors qu'en réalité il pleuvait
- ✓ La classe beau temps déduite par le modèle contient 23 observations bien classées et 7 (6+1) mauvaises .

Également, à partir de cette matrice, on peut calculer des mesures très importantes :

## ▪ Taux d'erreur :

correspond à la qualité générale du modèle. Pour cela, on va diviser les bonnes prédictions (somme de la diagonale) par le nombre total de prédiction, le tout soustrait de 1. Dans notre exemple, on a une précision égale à 29%.

## ▪ Précision par classe :

correspond à la qualité de la classe. On divise le nombre d'éléments bien classés dans la classe par le nombre total d'éléments attribués à la classe. Par exemple, il y a une précision de 74% (31/42) dans la classe Pluie [18].

▪ **Rappel:** On va diviser les bonnes prédictions (somme de la diagonale) par le nombre total de prédiction \*100.

## 3-État de l'art sur quelques techniques de segmentation d'images IRM

Sur l'article d'Abdelouahab Moussaoui et al [19], Ils ont proposé une nouvelle architecture de fusion de données basée sur la théorie possibiliste pour la segmentation des images médicales.

Le processus de fusion est décomposé en trois phases fondamentales. Ils modélisent tout d'abord les informations dans un cadre théorique commun. Le formalisme retenu consiste à faire la coopération entre l'algorithme FCM (*Fuzzy C-Means*) dont la contrainte d'appartenance d'un individu à une classe est gérée d'une manière relative et l'algorithme possibiliste PCM (*Possibilistic C-means*) pour les points aberrants. Ils agrègent ensuite ces différentes informations par un opérateur de fusion. Celui-ci doit affirmer les redondances, gérer les complémentarités et prendre en compte les conflits soulignant souvent la présence d'une pathologie. Ils construisent enfin une information synthétique permettant d'exploiter les résultats de la fusion.

## ▪ les images utilisées :

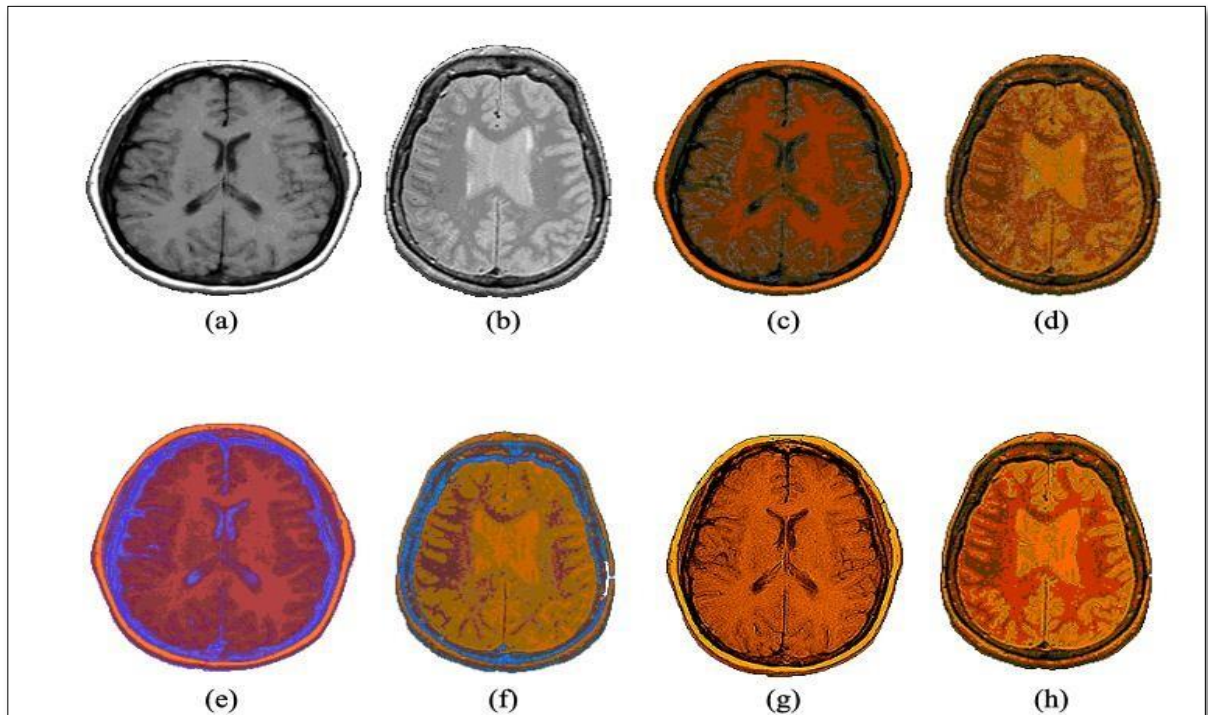
Des images réelles sur lesquelles ils ont travaillé, ces images ont été acquises dans le cadre de la collaboration entre le laboratoire LSI (Laboratoire Systèmes Intelligents : équipe image et

## CHAPITRE 1 : Segmentation d'Images médicale

signaux) de l'université Ferhat Abbas de Sétif et le l'hôpital universitaire de la ville de Sétif. Il s'agit d'images pondérées en T1, T2 et en densité de proton (Dp) pour des patients de différents âges (taille pixel =1mm, taille de matrice 512 X 512). Les images sont en formatDICOM (Digital Imaging and Communications in Medecine).

### ▪ Résultats :

La figure 15 illustre les résultats obtenus après segmentation des différents tissus cérébraux d'images (a) et (b) pondérées en T1 et en T2 sur des coupes axiales. Les images (c),(d), (e) et (f) sont les résultats de la segmentation par FCM et PCM successivement. Dans ces images (c, d, e, f) la distinction entre les différentes classes segmentées ne s'exprime pas complètement (problèmes d'initialisations des centres des classes), contrairement aux images (g), (h) issues de la segmentation par le système développé où on remarque la distinction entre les 3 classes du cerveau et la visibilité de la classe LCS, (invisible dans les deux approches précédentes).



**Figure I.16:** Résultats de segmentation

### ▪ Discussion et Critiques

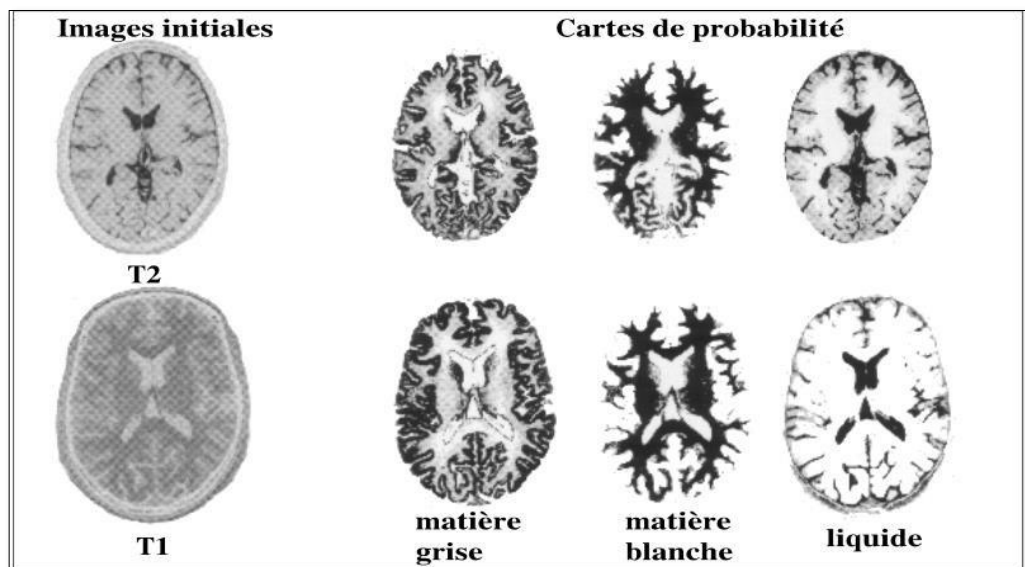
Ce travail a opté pour les méthodes d'apprentissages non supervisé FCM, PCM, et leur hybridation le résultat montre que les images issues de l'approche proposée et le PCM caractérisent bien les différentes classes cérébrales MB, MG, LCR. La zone entourée par le contour bleu représente une nouvelle classe non détectée par FCM et PCM. Aussi il ne faut oublier que Le coefficient de flou peut avoir une influence sur la performance de *clustering* de FCM. Sans oublier

# CHAPITRE 1 : Segmentation d'Images médicale

le manque d'une connaissance a priori qui aide beaucoup à améliorer les résultats.

L'approche que nous décrivons ci-dessous représente une méthode utilisée en classification bayésienne. En effet, Friston [20] propose un moyen d'introduire de la connaissance a priori dans le processus de segmentation.

Friston [20] exploitent le principe de la classification bayésienne pour segmenter des images 3D du cerveau en 4 classes (matière grise, matière blanche, liquide céphalo-rachidien et autres). L'originalité de l'approche réside dans l'introduction de connaissance a priori pour initialiser les cartes de probabilité. La connaissance a priori est constituée par un atlas probabiliste représentant la probabilité moyenne pour un pixel d'appartenir à une classe donnée. Ces cartes ont été établies à partir d'un grand nombre d'images segmentées fournies par le Montreal Neurological Institute. L'algorithme est itératif et fondé sur le calcul des paramètres des fonctions gaussiennes représentant les distributions.



**Figure I.17: Segmentation probabiliste en 3 classes de 2 types de séquences d'acquisition.**

Une limite importante de cette approche est qu'elle est fondée sur une hypothèse d'uniformité des niveaux de gris pour un même tissu dans toutes les zones de l'image, hypothèse non vérifiée dans la réalité. Par conséquent, des erreurs de classification peuvent résulter du traitement choisi. En contrepartie, l'introduction de connaissance a priori sous forme de probabilité permet de simplifier les calculs pour la classification et d'éviter le recours à un processus d'isolement du cerveau par rapport au reste de l'image.

L'introduction de connaissance a priori au sein d'un processus de segmentation. Cela permet de mieux contraindre un problème tout en tirant parti de résultats acquis antérieurement.

# CHAPITRE 1 : Segmentation d'Images médicale

## 4-Quelques difficultés de la segmentation des images médicales

### + Bruit :

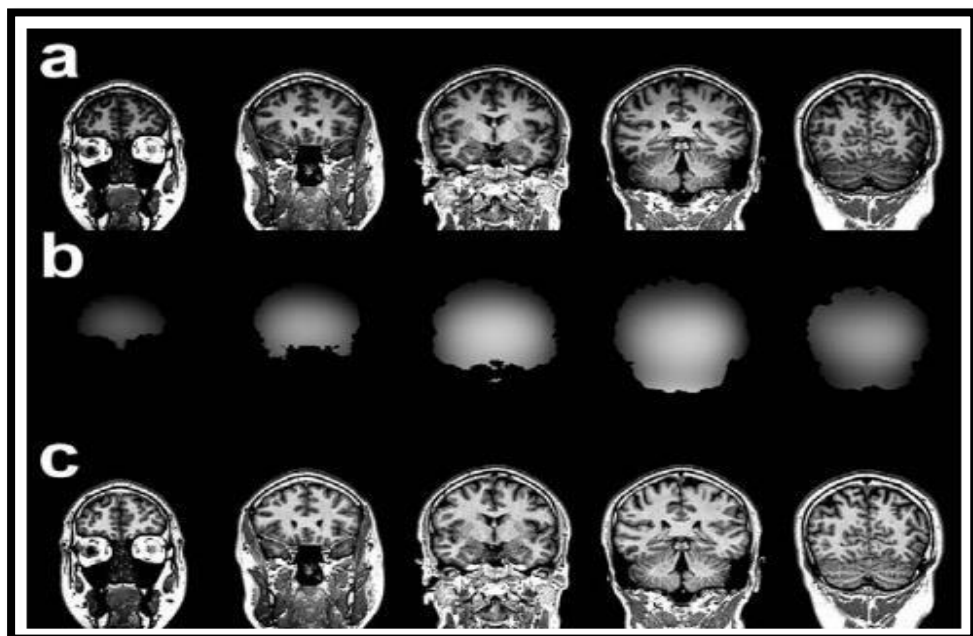
Le bruit d'image caractérise la présence d'informations parasites qui s'ajoutent de façon aléatoire aux détails dans les images acquises. Il est plus particulièrement visible dans les zones peu éclairées, où le rapport signal/bruit est faible. Il a pour conséquence la perte de netteté dans les détails [8].

### + Inhomogénéités du signal dans le volume :

il est techniquement très difficile de maintenir un champ magnétique spatialement constant. Le champ magnétique radiofréquence  $\vec{B}_1$  d'intensité  $B_1$  est en réalité sujet à de lentes variations spatiales. Ces variations contribuent à la variabilité de l'intensité du signal au sein d'un même tissu au-delà de sa variabilité naturelle (Figure17).

C'est un problème majeur car il perturbe la lecture de l'image pour le radiologue mais modifie aussi localement les statistiques de signal sur lesquelles sont basées les algorithmes de segmentation cérébrale, l'hypothèse sous-jacente de ces algorithmes est qu'un tissu est représenté par des intensités similaires dans l'image.

L'inhomogénéité du champ  $B_1$  varie en fonction des machines et augmente avec la fréquence et donc l'intensité du champ magnétique. C'est donc un problème central étant donné l'accroissement régulier de l'intensité du champ magnétique utilisé en IRM [19].



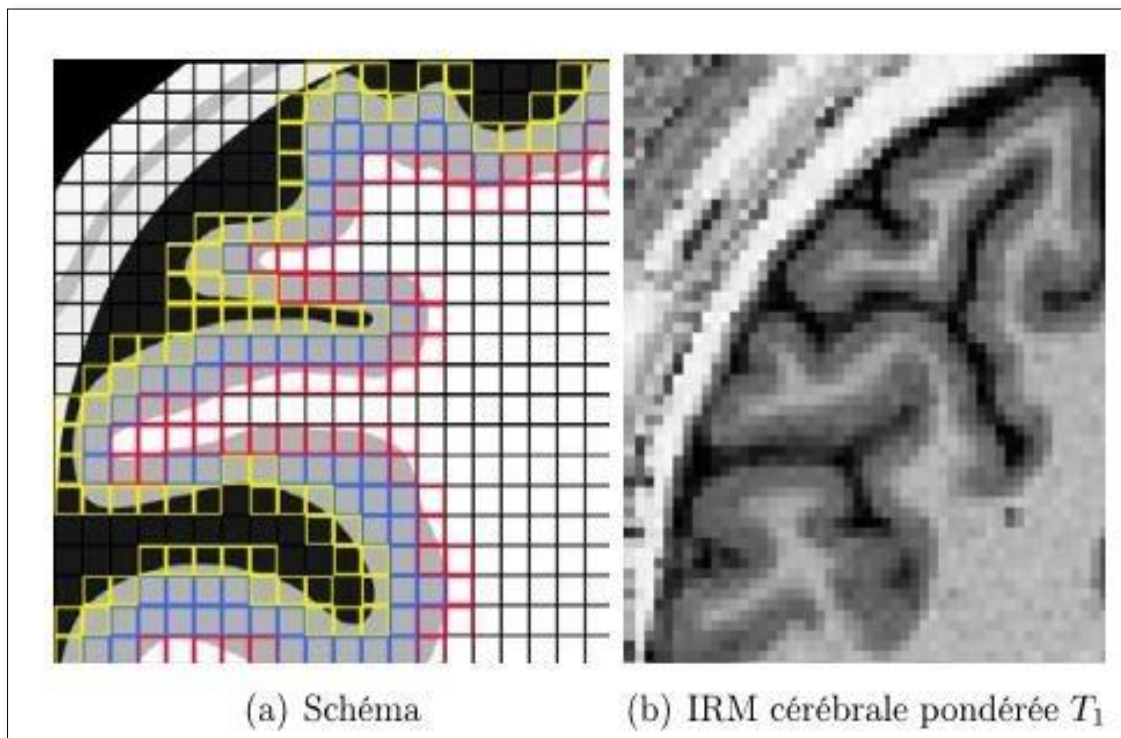
**Figure I.18:** Exemple d'image pondérée en T affectée par l'artefact d'inhomogénéité de champ magnétique radiofréquence. (a) Image originelle. (b) Carte d'estimation de l'inhomogénéité de champ magnétique B. (c) Image corrigée.

# CHAPITRE 1 : Segmentation d'Images médicale

## ✚ Effets de volume partiel

En IRM, le signal obtenu est discrétisé spatialement, c'est-à-dire que l'objet d'intérêt est imagé à travers une grille bi ou tridimensionnelle lors de l'encodage spatial. L'élément unitaire d'une image est appelée pixel ou voxel selon que l'image soit 2D ou 3D. À chaque voxel est associée une unique valeur du signal. Cependant, un voxel peut être situé à la frontière entre deux tissus et le signal associé à cette position reflète alors les phénomènes de résonance magnétique nucléaire de deux entités aux propriétés tissulaires différentes.

Ce phénomène est appelé "effet du volume partiel" et est un artefact inhérent aux systèmes d'imagerie numérique. La majeure partie des algorithmes de traitement d'images IRM prend en compte ces effets[19].



**Figure I.19:** (a) Schéma représentatif en 2D d'une structure cérébrale imagée à travers une grille de pixels. Les pixels rouges et jaunes sont sujets aux effets de volume partiel.

Les bleus contiennent uniquement de la substance grise. (b) Exemple d'une région d'intérêt d'une image obtenue avec une séquence pondérée en T.

## ✚ Artéfact de mouvement

L'artéfact de mouvement est un des artéfacts les plus fréquemment rencontré. Il se constitue

# CHAPITRE 1 : Segmentation d'Images médicale

lorsqu'il y a translation dans l'espace du segment étudié au cours de l'acquisition. Il ya deux types de mouvements rencontrés [8]:

- ❖ Les mouvements périodiques: Ce sont les mouvements de la respiration, les battements cardiaques et les flux sanguins.
- ❖ Les mouvements apériodiques: Ce sont les mouvements du patient, les mouvements oculaires, la déglutition, le péristaltisme digestif et le flux du liquide cérébro-spinal.

Ils ont pour conséquence la dispersion du signal : image floue de la structure en mouvement .

## **5-Conclusion**

Dans ce chapitre nous avons présenté l'imagerie médicale IRM et les méthodes les plus utilisée pour la segmentation. Nous avons vu que la segmentation par classification permet de regrouper des objets en classes. Par contre, les méthodes (croissance de régions et division- fusion) sont utilisées pour la segmentation des images, en regroupant les objets suivant les critères d'homogénéités, On a vu aussi les techniques de la construction d'un ensemble de classifieurs en se basant sur les méthodes d'apprentissage supervisés (connaissance a priori). Ensuite on a citez quelque travaux concernant la segmentation d'images IRM .Et enfin nous avons discuté sur quelques difficultés de la segmentation des images médicales.

# CHAPITRE 02 :

## Apprentissage Automatique

# CHAPITRE 2 : Apprentissage Automatique

## 1-Introduction :

Les quantités d'informations collectées chaque jour dans le monde ne cessent d'augmenter. Il est devenu urgent et vital de recourir à des techniques et méthodes pour extraire de l'information à partir de cette masse volumineuse de données, ce qui marque le début d'une nouvelle ère informatique notamment dans le domaine de l'apprentissage automatique "Machine Learning".

Dans ce chapitre, nous allons d'abord faire un tour d'horizon sur l'apprentissage automatique le Data Mining et ses outils, ensuite nous verrons les principaux algorithmes de l'apprentissage automatique, à la fin de ce chapitre introductif, nous parlerons de les méthodes d'agrégation Bagging et Adaboost.

## 2-L'apprentissage automatique

### 2- 1- Définition :

Le Machine Learning ou « apprentissage automatique » en français est un concept qui fait de plus en plus parler de lui dans le monde de l'informatique. Il se rapporte au domaine de l'intelligence artificielle. Encore appelé « apprentissage statistique », ce terme renvoie à un processus de développement, d'analyse et d'implémentation conduisant à la mise en place de procédés systématiques. Pour faire simple, il s'agit d'une sorte de programme permettant à un ordinateur ou à une machine un apprentissage automatisé, de façon à pouvoir réaliser un certain nombre d'opérations très complexes.

L'objectif visé est de rendre la machine ou l'ordinateur capable d'apporter des solutions à des problèmes compliqués, par le traitement d'une quantité astronomique d'informations. Cela offre ainsi une possibilité d'analyser et de mettre en évidence les corrélations qui existent entre deux ou plusieurs situations données, et de prédire leurs différentes implications[21].

### 2-2 Modélisation

L'apprentissage automatique d'une machine toujours concerne un ensemble de tâches concrètes T. Pour déterminer la performance de la machine, on utilise une mesure de la performance P. La machine peut avoir à l'avance un ensemble d'expérience E ou elle va enrichir cet ensemble plus tard.

## CHAPITRE 2 : Apprentissage Automatique



Figure II.01: Schéma de modélisation d'une machine d'apprentissage

Donc, l'apprentissage automatique pour la machine est qu'avec l'ensemble de tâches T que la machine doit réaliser, elle utilise l'ensemble d'expériences E telle que sa performance sur T est améliorée [22].

### 3 - Domaines de l'apprentissage automatique :

Les principaux domaines de l'apprentissage automatique sont les fouilles de données et l'intelligence artificielle :

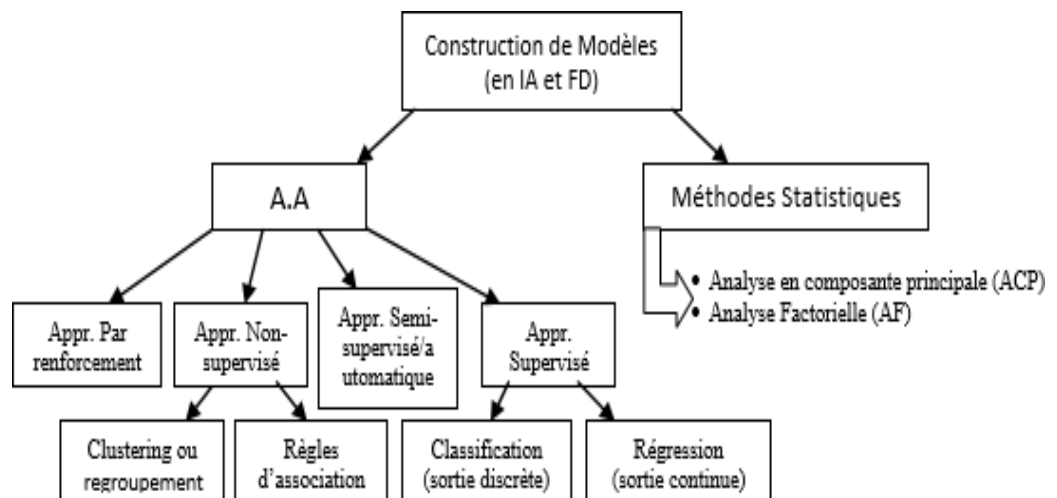


Figure II.02 : Schéma des différentes techniques issues de l'IA et FD pour la construction de modèles de données

## CHAPITRE 2 : Apprentissage Automatique

### 3-1- La fouille de données et ses méthodes :

La fouille de données apparaît au milieu des années 1990 aux États-Unis comme une nouvelle discipline à l'interface de la statistique et des technologies de l'information , bases de données, intelligence artificielle, apprentissage automatique (« machine Learning ») [23].

Le data Mining est l'ensemble des méthodes scientifiques destinées à l'exploration et l'analyse de grandes bases de données informatiques en vue de détecter dans ces données des profils-type, des comportements récurrents, des règles, des liens, des tendances inconnues (non fixées a priori), des structures particulières restituant de façon concise l'essentiel de l'information utile pour l'aide à la décision[24].

Le processus du data Mining "fouille de données" représente une étape d'un processus plus grand : l'ECD. D'après [25], un processus d'ECD est constitué de quatre phases qui sont : le nettoyage et intégration des données, le prétraitement des données, la fouille de données et enfin l'évaluation et la présentation des connaissances

La figure 03 récapitule ces différentes phases ainsi que les enchaînements possibles entre ces phases.

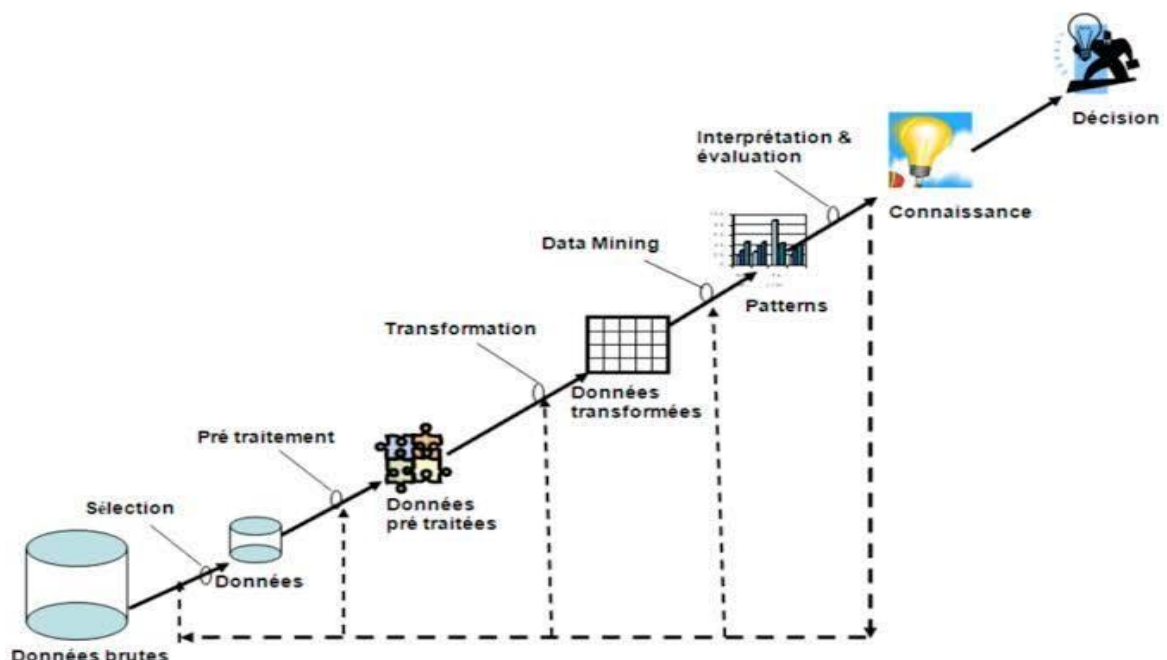


Figure II.03 : Processus d'extraction de connaissances à partir des données

## CHAPITRE 2 : Apprentissage Automatique

Pour arriver à exploiter ces quantités importantes de données, le data Manning utilise des méthodes d'apprentissages automatiques. Ses méthodes sont de deux types : les méthodes descriptives et les méthodes prédictives.

### a ) Méthodes descriptives :

Le principe de ces méthodes est de pouvoir mettre en évidence les informations présentes dans le data warehouse mais qui sont masquées par la masse de donnée [26]. Parmi les techniques et algorithmes utilisés dans l'analyse descriptive, on cite :

#### ➤ La segmentation

La segmentation est une tâche d'apprentissage car on ne dispose d'aucune autre information préalable que la description des exemples. [27].

#### ➤ Les règles d'association

Les règles d'association sont traditionnellement liées au secteur de la distribution car leur principale application est " l'analyse du panier de la ménagère " qui consiste en la recherche d'associations entre produits sur les tickets de caisse. Le but de la méthode est l'étude de ce que les clients achètent pour obtenir des informations sur qui sont les clients et pourquoi ils font certains achats. La méthode recherche quels produits tendent à être achetés ensemble [27].

### b) Méthode prédictives :

Contrairement à l'analyse descriptive, cette technique fait appels à de l'intelligence artificielle. L'analyse prédictive, est comme son nom l'indique une technique qui va essayer de prévoir une évolution des événements en se basant sur l'exploitation de données stockés dans le data warehouse.

#### ➤ La classification

La classification consiste à examiner des caractéristiques d'un élément nouvellement présenté afin de l'affecter à une classe d'un ensemble prédéfini. Par exemple classification d'un individu dans une classe parmi deux classes existantes : client dépensif ou client à budget[26].

- **Scoring**

Prédiction de valeur inconnu à partir de valeur connu .exemple: accordé oui ou non un prêta un client [26].

## CHAPITRE 2 : Apprentissage Automatique

### 3-2- L'intelligence Artificielle :

L'intelligence artificielle (aussi appelée intelligence informatique) est définie comme "l'intelligence des machines et des logiciels". Elle s'appuie sur l'hypothèse selon laquelle il est possible de décrire l'intelligence des êtres humains et, pour des machines ou des logiciels, de la simuler. Ces machines ou logiciels sont capables de raisonner, planifier, apprendre, percevoir et traiter les informations comme l'esprit humain et donc de faciliter l'existence des humains. Ils peuvent penser et agir à notre place, par exemple nettoyer, déplacer des objets, effectuer des diagnostics médicaux, dispenser des soins écrire du code logiciel, établir les meilleures prévisions météorologiques composer de la musique et jouer avec nous. L'intelligence artificielle est un champ d'étude interdisciplinaire réunissant l'informatique, les neurosciences, la psychologie, la linguistique et la philosophie [28].

L'apprentissage automatique ou apprentissage statistique, champ d'étude de l'intelligence artificielle, concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine (au sens large) d'évoluer par un processus systématique, et ainsi de remplir des tâches difficiles ou problématiques à remplir par des moyens algorithmiques plus classiques.

### 4 -Types de l'apprentissage automatique

Il y a différents types d'apprentissage utilisés en apprentissage automatique. Nous verrons l'apprentissage supervisé, non supervisé, semi-supervisé et par renforcement. Chaque type d'apprentissage a leur spécificité et son utilisation en fonction du problème que nous rencontrons..

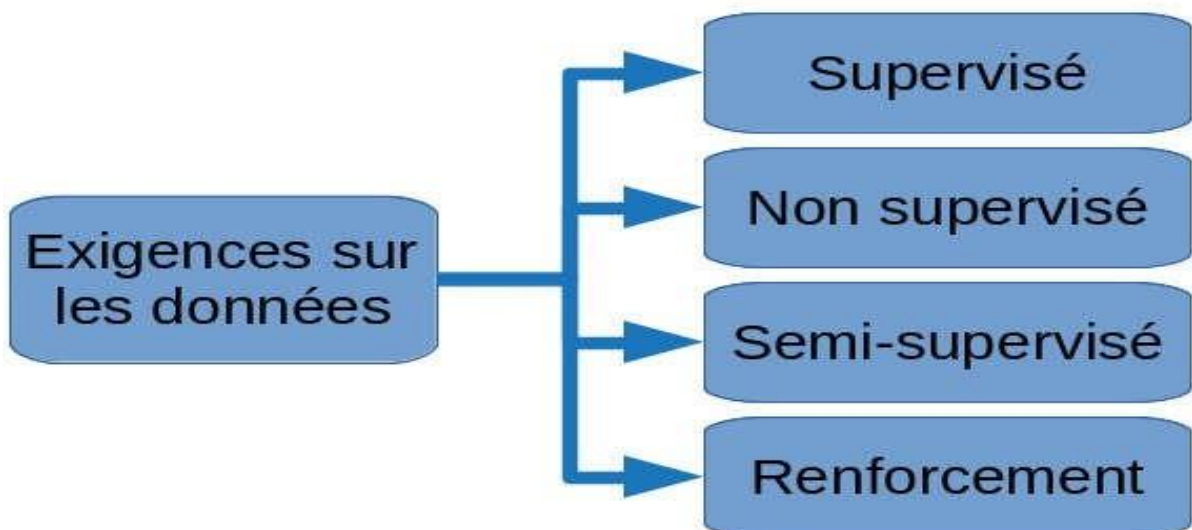


Figure II.04: Les différents types d'apprentissage automatique

## CHAPITRE 2 : Apprentissage Automatique

### 4-1- Apprentissage supervisé :

L'apprentissage supervisé ou *supervised learning* est une approche d'apprentissage où nous connaissons les données sur lesquels nous travaillons. Par exemple, dans le cas d'un classificateur d'images, nous allons avoir un ensemble de données labellisées. Ainsi, nous connaissons l'étiquette de chaque image. À partir de ces étiquettes, nous allons pouvoir apprendre à notre modèle les différentes catégories représentées. Ainsi, à chaque prédiction de notre modèle, nous pouvons lui dire s'il a donné une bonne prédiction ou s'il s'est trompé. De ce fait, notre système apprend de ses erreurs. Il va chercher à minimiser l'erreur en fonction des données que nous allons lui donner.

Généralement, dans un apprentissage supervisé, nous avons trois grands temps. Un temps d'apprentissage, où nous allons donner à notre modèle un ensemble de données, ou il pourra s'entraîner. Puis une phase de validation, où nous allons chercher à vérifier l'apprentissage de notre système. Pour ce faire, nous allons prendre un ensemble de données que nous n'avons jamais montré à notre modèle. Puis, nous allons chercher à vérifier si notre modèle arrive à bien différencier les différentes catégories. Enfin, nous avons une phase de production où nous allons pouvoir utiliser notre système [29].

L'apprentissage supervisé est utilisé dans les applications financières pour l'évaluation de crédit, dans les applications biologiques pour la détection des tumeurs et la découverte de médicaments, dans les applications énergétiques pour la prévision des prix et de la consommation, et dans les applications de reconnaissance des formes pour la parole et les images.



Figure II.05 : Schéma d'un modèle supervisé

## CHAPITRE 2 : Apprentissage Automatique

### 4-2. Apprentissage non supervisé :

Dans un apprentissage non supervisé ou un supervised learning, nous ne connaissons pas la catégorie d'appartenance de nos données. Contrairement à l'apprentissage supervisé, nous n'avons ici pas de label sur nos données. Ainsi, l'objectif va être, pour notre modèle, de repérer des similarités et, en fonction de celle-ci, de déterminer certaines données qui appartiennent à la même catégorie.

Pour illustrer ce type d'apprentissage, nous avons un ensemble de données d'une population d'animaux. Nous avons récolté un ensemble de caractéristiques comme leur âge, leur poids, leur taille, L'objectif va être de déterminer si certains animaux peuvent appartenir à la même espèce et ainsi les regrouper ensemble.

Nous mesurons la qualité d'une méthode de classification non supervisée par la capacité à notre modèle de découvrir certains motifs cachés [29].

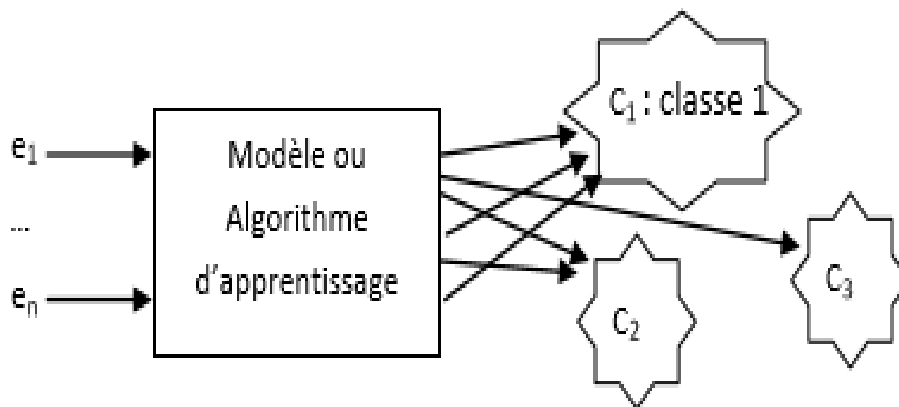


Figure II.06: Schéma d'un modèle non supervisé

### 4-3- Apprentissage semi-supervisé :

L'apprentissage semi-supervisé ou semi-supervised learning se situe entre le supervisé et le non-supervisé. En effet, dans ce type d'apprentissage, nous avons un ensemble de données étiquetées et non étiquetées. Ce qui est intéressant avec cette méthode est que nous laissons une certaine liberté à notre modèle. En effet, nous lui donnons des données sur lesquelles il va pouvoir extraire différentes caractéristiques et sur d'autres, il va devoir trouver des caractéristiques qui lui permettent de s'améliorer.

Ce type d'apprentissage est intéressant lorsque nous avons énormément de données. En effet, cela peut être fastidieux d'étiqueter l'intégralité de nos données. Cette méthode d'apprentissage nous permet d'éviter ce problème [29].

## CHAPITRE 2 : Apprentissage Automatique



Figure II.07: Schéma d'un modèle semi-supervisé

### 4-4 Apprentissage par renforcement :

Dans le cas de l'apprentissage par renforcement, nous allons avoir un agent présent dans un environnement où il sera amené à prendre des décisions. Notre agent va se retrouver dans une série d'états où il aura à sa disposition un ensemble d'actions. L'objectif de notre agent est de prendre la meilleure décision possible. Pour définir la meilleure action, nous allons attribuer à chaque action une récompense positive ou négative. Ainsi, l'objectif pour notre agent est d'obtenir la meilleure récompense possible.

Pour illustrer ce type d'apprentissage, nous pouvons prendre le cas d'un aspirateur autonome. L'environnement dans lequel il se situe est la pièce et son objectif est de la nettoyer.

Pour ce faire, on peut attribuer pour chaque position de la pièce une récompense lui indiquant qu'il n'a pas encore nettoyé cette partie. Une fois passé sur une position, nous lui attribuons une récompense négative lui indiquant qu'il n'est pas censé repasser sur la même position. On pourrait aussi imaginer que lorsque notre robot aspirateur passe de l'état en charge à l'état à besoin d'être chargé, il faut qu'il privilégie le chemin de retour à la borne de chargement [29].

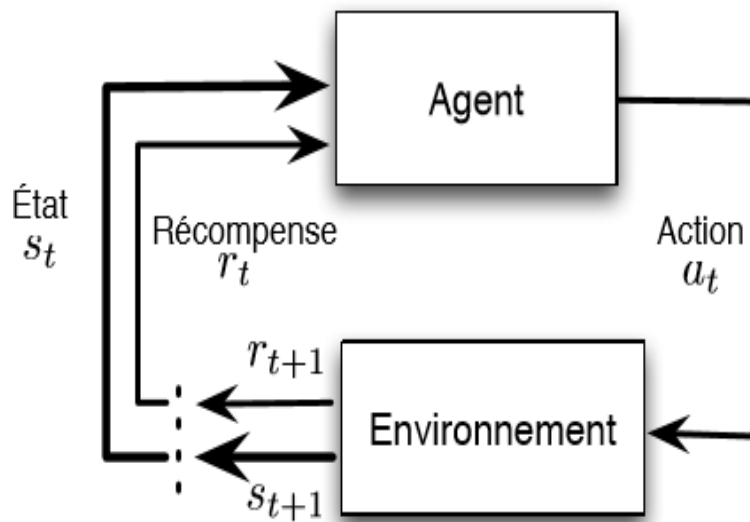


Figure II.08: Schéma d'un modèle par renforcement

# CHAPITRE 2 : Apprentissage Automatique

## 5- Les principaux algorithmes d'apprentissage automatique

On va s'intéresser sur les algorithmes d'apprentissage supervisé pour l'apprentissage automatique. Il existe deux types paramétrique et non paramétrique.

### 5-1- Non paramétrique :

Les approches dites non paramétriques (classification hiérarchique, méthode des centres mobiles) basée sur l'hypothèse : plus deux individus sont proches, plus ils ont de chances de faire partie de la même classe. En plus ce que distingue cette approche est qu'on ne fait aucune hypothèses sur le modèle que suivent les données. C'est le cas des plus proches voisins (k-PPV), arbre de décision, le boosting [30].

#### 5-1-1 -Les k plus proches voisins :

La méthode des "k plus proches voisins" fait partie des méthodes les plus simples d'apprentissage supervisé pouvant être utilisée pour les cas de régression et de classification.

Nous allons nous intéresser à la théorie derrière cet algorithme dans le cas de la classification supervisée et son implémentation sur les logiciels R et Python-

Les 'k plus proches voisins' ou k-nearest neighbors en anglais (d'où l'appellation knn) est une méthode non paramétrique dans laquelle le modèle mémorise les observations de l'ensemble d'apprentissage pour la classification des données de l'ensemble de test.

En effet, cet algorithme est qualifiée comme paresseux (Lazy Learning) car il n'apprend rien pendant la phase d'entraînement. Pour prédire la classe d'une nouvelle donnée d'entrée, il va chercher ses K voisins les plus proches (en utilisant la distance euclidienne, ou autres) et choisira la classe des voisins majoritaires.

pour appliquer cette méthode, les étapes à suivre sont les suivantes :

- On fixe le nombre de voisins k.
- On détecte les k-voisins les plus proches des nouvelles données d'entrée quel'on veut classer.
- On attribue les classes correspondantes par vote majoritaire.

Mais, comment choisit-on ce paramètre k lors de l'implémentation de l'algorithme ?

- On fait varier k
- Pour chaque valeur de k, on calcule le taux d'erreur de l'ensemble de test.

## CHAPITRE 2 : Apprentissage Automatique

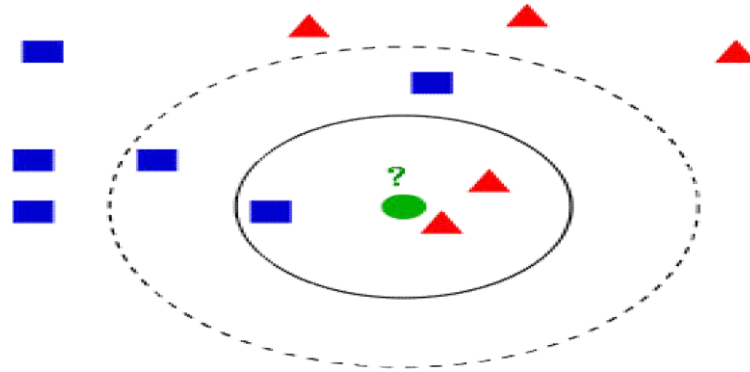


Figure II.09 : Fonctionnement de k plus proche voisins

### Avantage :

- Facile à comprendre
- Apprentissage rapide

### Inconvénients :

- Pas efficace pour des jeux de données larges.
- L'estimation de ce modèle devient de mauvaise qualité quand le nombre de variables explicatives est grand [31].

### 5-1-2 Les arbres de décision :

La méthode des arbres de décision est l'une des plus intuitives et des plus populaires, elle fournit des règles explicites de classement et supporte bien les données hétérogènes, manquantes et les effets non linéaires. Pour les applications relevant du marketing de bases de données, actuellement la seule grande concurrente de l'arbre de décision est la régression logistique, cette méthode étant préférée dans la prédiction du risque en raison de sa plus grande robustesse. Remarquons que les arbres de décision sont à la frontière entre les méthodes prédictives et descriptives, puisque leur classement s'opère en segmentant la population à laquelle ils s'appliquent : ils ressortissent donc à la catégorie des classifications hiérarchiques descendantes supervisées.

Il existe deux principaux types d'arbre de décision en fouille de données :

- Les arbres de classification : (*Classification Tree*) permettent de prédire à quelle classe la variable-cible appartient, dans ce cas la prédiction est une étiquette de classe.
- Les arbres de régression : (*Regression Tree*) permettent de prédire une quantité réelle (par exemple, le prix d'une maison ou la durée de séjour d'un patient dans un hôpital), dans ce cas la prédiction est une valeur numérique.

## CHAPITRE 2 : Apprentissage Automatique

Les arbres utilisés dans le cas de la régression et dans le cas de la classification présentent des similarités mais aussi des différences, en particulier en ce qui concerne la procédure utilisée pour déterminer les séparations des branches [32].

### Avantages

- Bonne lisibilité du résultat
- Traitement de tout type de données : l'algorithme peut prendre en compte tous.
- Il est robuste au bruit.
- Méthode extensible et modifiable : la méthode peut être adaptée pour résoudre des tâches d'estimation et de prédiction.

### Inconvénient

- Méthode sensible au nombre de classes

### 5-2-Paramétrique :

La seconde grande famille des méthodes de classification, concerne les approches probabilistes qui utilisent une hypothèse sur la distribution des individus à classer, c'est-à-dire, que l'on suppose que l'on connaît la forme du modèle qui a généré les données. Par exemple, on peut considérer que les individus de chacune des classes suivent une loi normale. Le problème qui se pose, est de savoir déterminer ou estimer les paramètres des lois (moyenne, variance) et à quelle classe les individus ont le plus de chances d'appartenir à partir de l'ensemble d'apprentissage. Les paramètres d'une loi peuvent être déterminés de maintes façons, C'est le cas par exemple des classifications bayésiennes ou des machines à vecteurs de support SVM [33].

#### 5-2-1-Classification naïve bayésienne :

La méthode de classification naïve bayésienne est un algorithme d'apprentissage supervisé (*supervised machine learning*) qui permet de classer un ensemble d'observations selon des règles déterminées par l'algorithme lui-même. Cet outil de classification doit dans un premier temps être entraîné sur un jeu de données d'apprentissage qui montre la classe attendue en fonction des entrées. Pendant la phase d'apprentissage, l'algorithme élabore ses règles de classification sur ce jeu de données, pour les appliquer dans un second temps à la classification d'un jeu de données de prédiction. Le classificateur bayésien naïf implique que les classes du jeu de données d'apprentissage soient connues et fournies, d'où le caractère supervisé de l'outil [34].

## CHAPITRE 2 : Apprentissage Automatique

### Avantage :

- Il est possible de proposer une connaissance « a priori » avant de construire une hypothèse.
- Optimal, minimise l'erreur théorique.
- Permet l'application de modèles de plus en plus complexes.

### Inconvénients :

- Approche subjective.
- Nécessité de définir une distribution a priori.
- Calculs coûteux en mémoire et en temps.

### 5-2-2- Les machines à vecteurs de support SVM :

Les SVM sont des algorithmes qui utilisent une transformation non linéaire des données d'apprentissage. Ils projettent les données d'apprentissage dans un espace de plus grande dimension que leur espace d'origine. Dans ce nouvel espace, ils cherchent l'hyperplan qui permet une séparation linéaire optimale des données d'apprentissage en utilisant les vecteurs de support et les marges définies par ces vecteurs.

La technique SVM fait partie des techniques classiques de fouille de données. Elle fait partie des méthodes d'apprentissage qui ont réalisé des performances meilleures que les méthodes statistiques traditionnelles en matière de classification. L'algorithme implémenté au niveau de l'outil Weka est la version SMO développée par John Platt (Platt, 1998). Il remplace toutes les valeurs manquantes et transforme les attributs nominaux en binaires. Par défaut, il normalise tous les attributs; ainsi la sortie est basée sur ces attributs normalisés et non pas sur les données originales. Il est utilisé seulement pour des problèmes de classification [22].

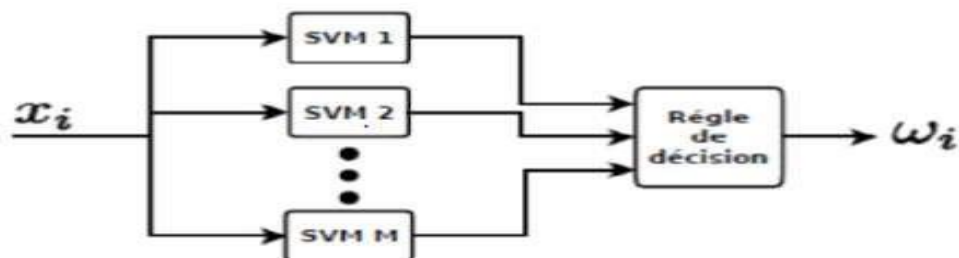


Figure II.10: Architecture parallèle de SVM

### Avantage :

- Capacité à traiter de grandes dimensionnalités.
- Traitement des problèmes non linéaires avec le choix des noyaux.

## CHAPITRE 2 : Apprentissage Automatique

### Inconvénients :

- Classification binaire.
- Modèle difficilement interprétable.

### 6- Classification non-supervisée par k-mens

Il s'agit d'un algorithme de classification non supervisée, ou nous ne disposons pas de données d'apprentissage, ou bien ces dernières ne sont pas étiquetées par des experts.

L'algorithme permet d'analyser un jeu de données caractérisées par un ensemble de descripteurs, afin de regrouper les données "similaires" en groupes (ou clusters).

La similarité entre deux données peut être inférée grâce à la "distance" séparant leurs descripteurs ; ainsi deux données très similaires sont deux données dont les descripteurs sont très proches. Cette définition permet de formuler le problème de partitionnement des données comme la recherche de K "données prototypes", autour desquelles peuvent être regroupées les autres données.

Après avoir initialisé ses centroïdes en prenant des données au hasard dans le jeu de données, K-means alterne plusieurs fois ces deux étapes pour optimiser les centroïdes et leurs groupes :

- Regrouper chaque objet autour du centroïde le plus proche.
- Replacer chaque centroïde selon la moyenne des descripteurs de son groupe.

### Avantage :

- Apprentissage non supervisée, pas nécessaire la présence d'experts.
- Performant et robuste.

### Inconvénients :

- Très sensible aux données d'apprentissage.

### 7- Les méthodes d'agrégation Boosting et Bagging :

#### 7-1- Boosting :

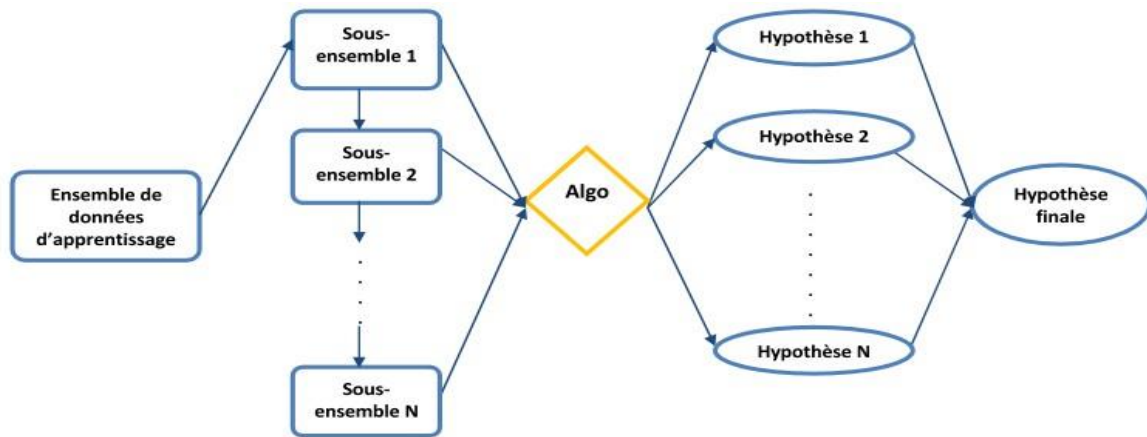
L'idée générale de Boosting est de développer de manière progressive l'ensemble des classifieurs, en ajoutant un seul classifieur à la fois. Chaque nouveau classifieur s'occupe des cas «mal classés» (c'est-à-dire les cas difficiles) par les autres classifieurs. La sélection des exemples d'apprentissage commence uniformément, et progresse en incrémentant la probabilité des exemples difficiles.

Le Boosting associe un poids à chaque exemple de l'ensemble d'apprentissage. Après chaque cycle d'apprentissage, les poids des exemples sont mis à jour en fonction de la performance du classifieur sur l'exemple correspondant. Dans la 1ère étape de l'algorithme tous les poids sont identiques, ensuite et pour chaque cycle, les poids des exemples bien classifiés seront décrémentés et

## CHAPITRE 2 : Apprentissage Automatique

les poids des exemples inexactement classifiés seront augmentés en obligeant le système de se concentrer aux exemples durs de l'ensemble d'apprentissage.

Un type très populaire du Boosting est l'AdaBoost (Le Boosting adaptatif), l'algorithme AdaBoost (expliqué dans le 1<sup>er</sup> chapitre) génère un ensemble de classifieurs et puis exécute un vote pondéré. La pondération de chaque classifieur est déterminée selon sa performance de prédiction. Les poids des exemples d'apprentissage sont mis à jour en se basant sur les classifieurs de l'étape précédente. Le but du Boosting est de forcer les classifieurs finaux de réduire au minimum l'erreur sur les différentes distributions des entrées. La classification finale est obtenue par le vote pondéré. Le Boosting a été appliqué avec succès à un grand nombre d'applications [35].

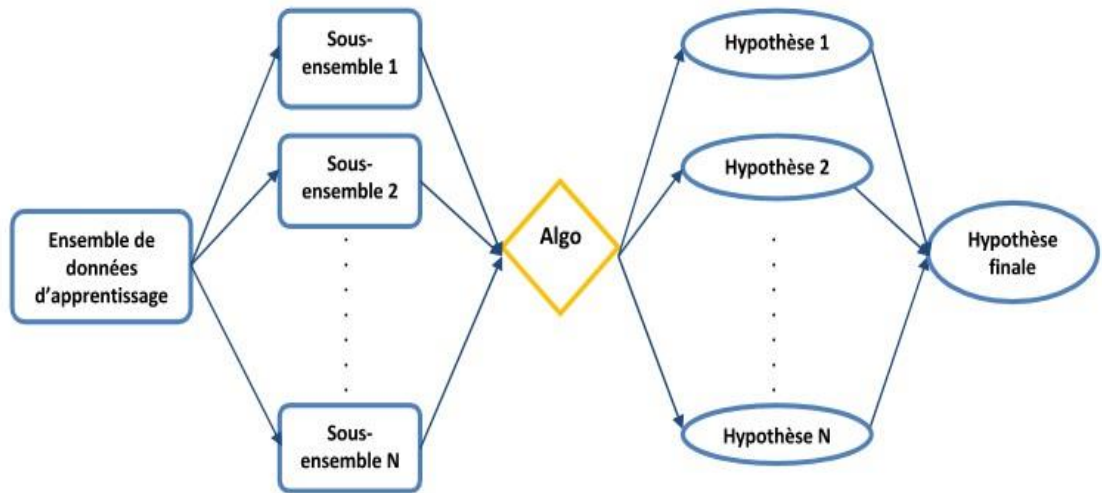


**Figure II.11: Fonctionnement du Boosting .**

### 7-2-Bagging :

Le Bagging est une méthode simple permettant de générer de multiples classifieurs et d'exécuter plusieurs sessions d'apprentissage avec le même classifieur sur différents sous-ensembles de l'ensemble d'apprentissage, ou avec un classifieur ayant des paramètres légèrement modifiés. Chaque session d'apprentissage crée un nouveau classifieur. Cette méthode engendre des ensembles d'apprentissage avec remplacement de l'ensemble d'apprentissage original, et chaque ensemble ayant comme résultat un classifieur légèrement différent après l'étape d'apprentissage. La technique utilisée pour la génération des sous-ensembles d'apprentissage est connue sous le nom "Bootstrapping". Le Bagging a montré de grandes capacités d'apprentissage sauf dans le cas des classifieurs faibles homogènes où il peut donner de légères améliorations. Dans ce cas, le Boosting est préférable [35].

## CHAPITRE 2 : Apprentissage Automatique



**Figure II.12: Fonctionnement du Bagging**

### **8-Conclusion :**

Dans ce chapitre nous avons présenté l'apprentissage automatique, et ses domaines : Intelligence artificielle et fouille de données, ainsi que les types d'apprentissage : supervisé, non supervisé, semi-supervisé et renforcement, et enfin nous avons cité les principaux algorithmes de l'apprentissage automatique. Pour ces algorithmes, nous les avons présenté avec suffisamment de détails et en présentant leurs avantages et inconvénients afin de bien pouvoir choisir le meilleur pour la classification des données IRM.

Dans le deuxième chapitre, nous présenterons le concept de l'imagerie médicale IRM, et nous étudierons la segmentation des images IRM ainsi que les différentes approches de segmentation utilisées

# CHAPITRE 03

## Moyennes Mobiles Locales Pour la Segmentation d'IRM Cérébrales

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

## 1. Introduction :

Dans ce troisième chapitre du mémoire, nous proposons une nouvelle méthode d'utilisation de l'algorithme des moyennes mobiles (k-means) pour la segmentation des IRM cérébrales. La technique consiste à appliquer l'algorithme k-means sur des sous-régions locales des coupes d'IRM (images 2D).

Dans la suite de ce chapitre, nous présentons d'abord l'algorithme k-means, puis le principe de la méthode proposée, et enfin, les différentes étapes de cette méthodes.

## 2. Algorithme k-means :

### 2.1- Définition :

L'algorithme des k-moyennes (ou K-means en anglais) est un algorithme de partitionnement de données relevant des statistiques et de l'apprentissage automatique (plus précisément de l'apprentissage non supervisé) [36], le plus connu et le plus utilisé, du fait de sa simplicité de mise en œuvre . C'est une méthode dont le but est de diviser des observations en K partitions (clusters) dans lesquelles chaque observation appartient à la partition avec la moyenne la plus proche [37]. Contrairement à d'autres méthodes dites hiérarchiques, qui créent une structure en « arbre de clusters » pour décrire les groupements [36], l'algorithme classique de K-means est le même que l'algorithme de quantification de Lloyd-Max. L'algorithme de regroupement autour de centres mobiles est généralement imputé à Forgy. [37]

En réalité, de nombreux travaux ont été menés parallèlement sur le thème des centres mobiles, introduisant des variantes. Cette méthode est connue, en anglais, sous le nom de k - moyennes (k-means). La méthode des k-moyennes est imputée par certains auteurs en France, dont l'algorithme commence également par un tirage des centres, mais contrairement à la technique des centres mobiles, chaque réaffectation d'un point entraîne une modification immédiate du centre correspondant.[37].

## CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

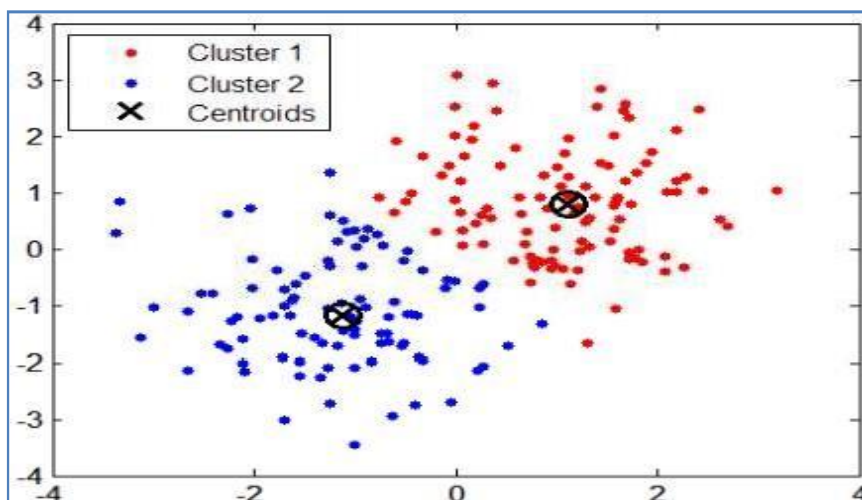


Figure III.1 : Classification à base de K-means

### 2-2- Principe de la méthode des K-Means:

Principe du K-means est de segmenter les données en k-groupes. Voici un exemple illustratif du déroulement de l'algorithme des K-Means dans la Figure 2 ci-dessous avec trois clusters.

Tout d'abord, k points sont choisis semi aléatoirement comme centre des clusters. Toutes les instances sont assignées au centre le plus proche d'eux, ceci étant calculé avec la distance euclidienne. [38]

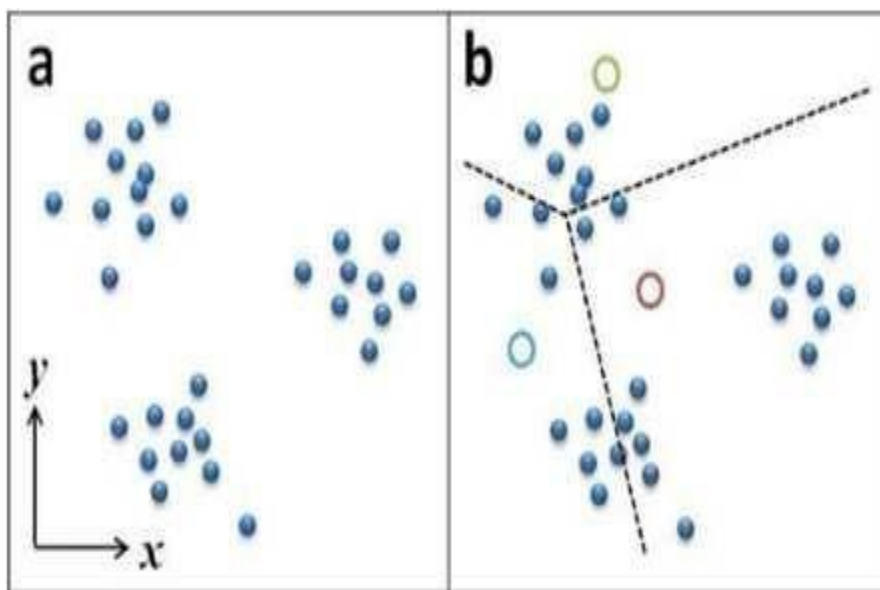
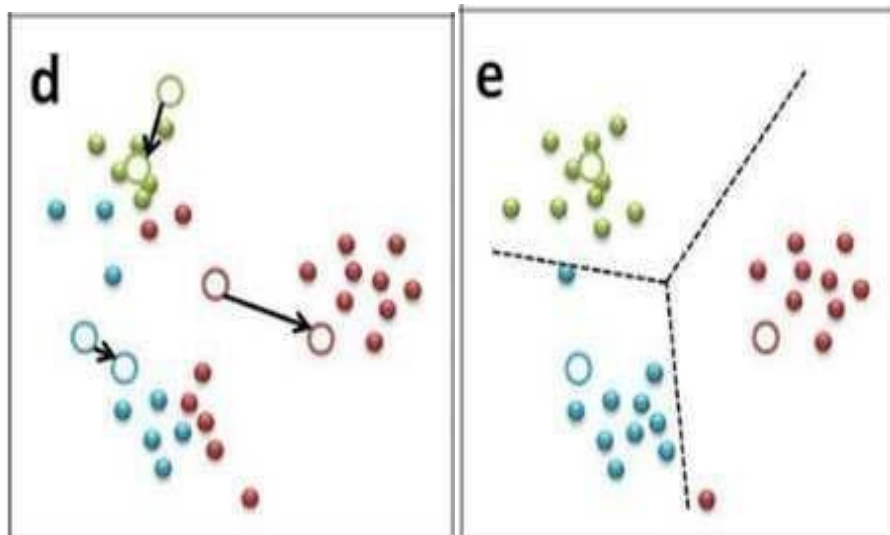


Figure III.2 : Exemple d'initialisation du K-Means avec trois clusters

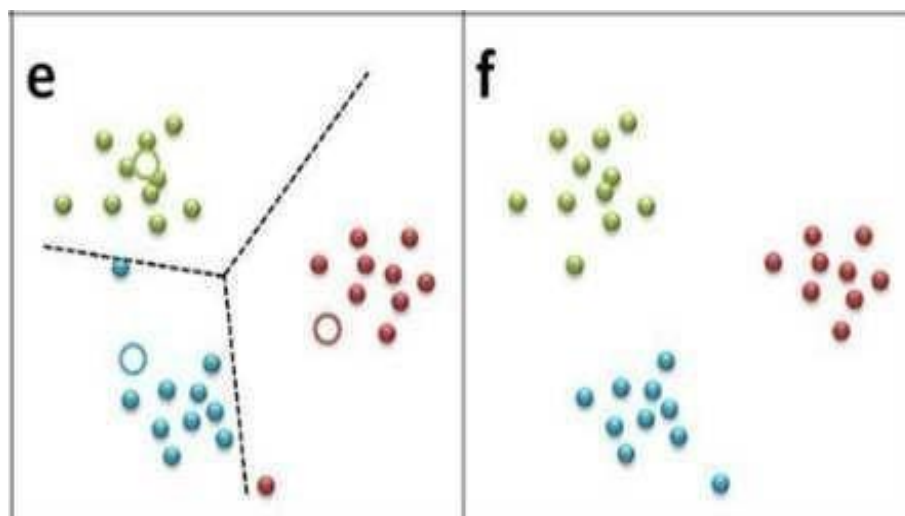
## CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

Ensuite, les centres de chaque cluster formés sont recalculés par rapport à la position des instances qu'ils contiennent. Puis, les instances sont réassignées à chacun des clusters en fonction de leur distance euclidienne par rapport aux nouveaux centres (Figure 3)



**Figure III.3 :** Mise à jour des centres après une itération de K-means

Enfin, ce processus est répété jusqu'au moment où les centres des clusters varieront très peu d'une itération à l'autre, on appelle ce changement la stabilisation des centres de gravité. Ainsi, on obtient des clusters qui ne se recoupent pas, et englobe toutes les instances du jeu de données (Comme le montre la Figure 4).



**Figure III.4 :** Exemples de clusters finaux obtenus avec K-means.

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

Le principe des K-means est de minimiser itérativement la fonction objective:

$$J(y, v) = \sum_{i=1}^N \sum_{k=1}^c \|y_i - v_k\|^2 \dots\dots\dots 1$$

- C : est le nombre de classe.
- N : est le nombre de données.
- $v_k$  :est le centre de classe i.
- $\| \cdot \|$  est la norme Euclidienne qui mesure la similarité entre un pixel est un autre.

Tenant compte de la nouvelle fonction objectif (1), la mise à jour des centres des classes v se fait avec l'équation(2):

$$v_k = \frac{1}{N} \sum_{i=1}^N y^i \dots\dots\dots 2$$

- n: nombres de données (individus ou pixels) de chaque classe.

## 2-3- Principe Algorithmique : [39]

L'algorithme suivant résume les étapes principales pour l'exécution :

**Entrée**  
 Ensemble de N données, noté par x  
 Nombre de groupes souhaité, noté par k

**Sortie**  
 Une partition de K groupes {C1,C2, Ck}

**Début**

1) **Initialisation aléatoire des centres Ck;**

**Répéter**

2) **Affectation :** générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche;

$$x_i \in C_k \text{ Si } \forall j |x_i - \mu_k| = \min |x_i - \mu_j|$$

Avec  $\mu_k$  le centre de la classe K ;

3) **Représentation :** Calculer les centres associe à la nouvelle partition;

$$\mu_k = \frac{1}{N} \sum_{x_i \in C_k} x_i$$

Jusqu'à convergence de l'algorithme vers une partition stable ;

Fin.

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

## 2-4- Organigramme de l'algorithme de k-means:

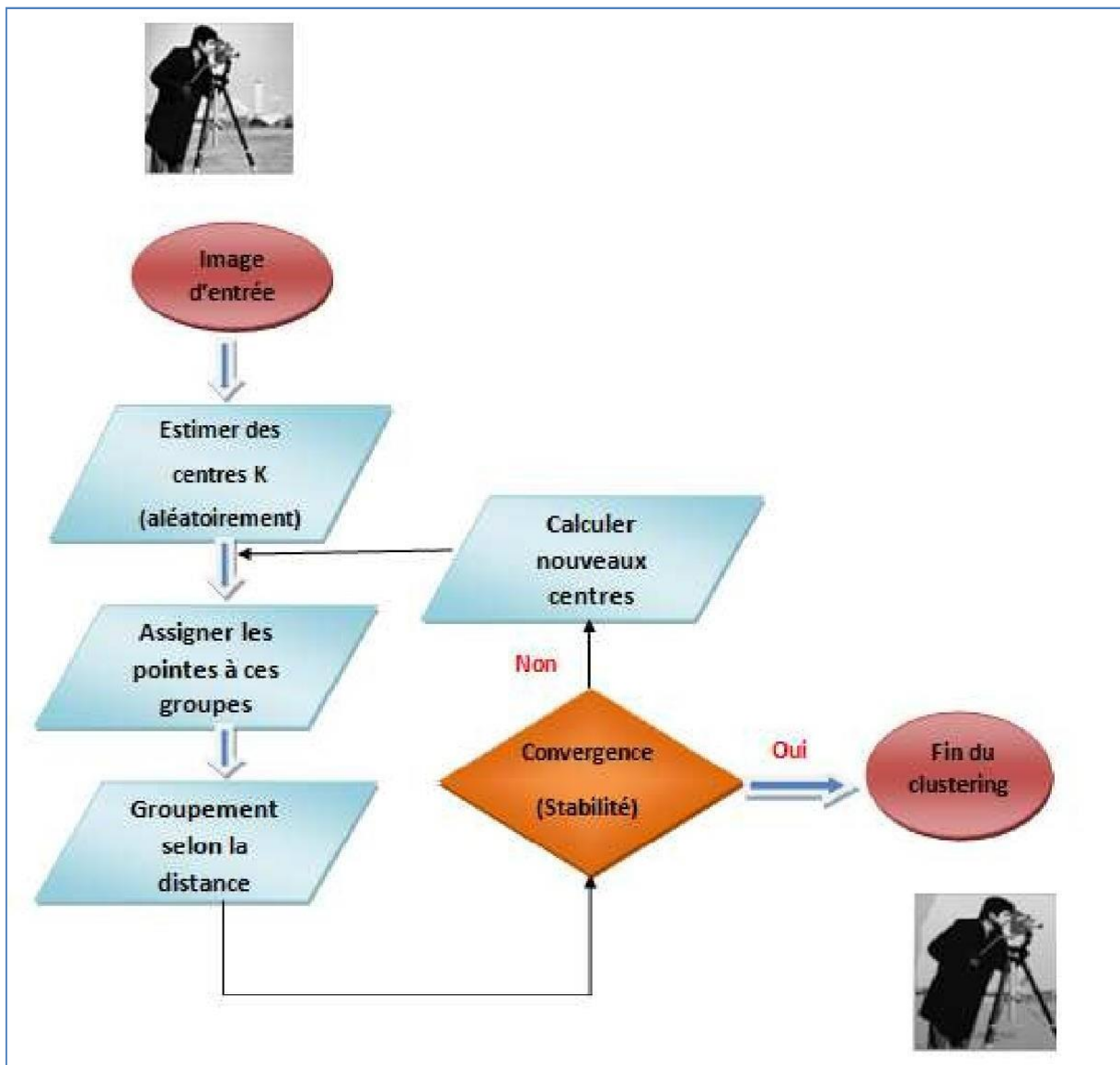


Figure III.5 : Organigramme de l'Algorithme K-Means

## 2-5-Convergence et initialisation des K-Means:

Convergence est définie comme un minimum local de l'énergie, qui se traduit par partition de l'espace des données en des classes séparées par des hyper plans. La qualité de la solution ainsi trouvée dépend fortement des noyaux initiaux. De plus la sensibilité de l'algorithme à l'initialisation est d'autant plus grande que la dimensionnalité des données est grande [40].

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

## 2-6- Choix du nombre K de classes (clusters):

L'algorithme classique des K-Means laisse un paramètre libre ; le nombre de classes, ce qui dans le cas de la segmentation d'images correspond au nombre d'intensités utilisées pour représenter l'image [40] .

## 2-7-Cas d'utilisation k-means :

K-Means en particulier et les algorithmes de clustering de façon générale ont tous un objectif commun : Regrouper des éléments similaires dans des clusters. Ces éléments peuvent être tous et n'importe quoi, du moment qu'ils sont encodés dans une matrice de données.

Les champs d'application de K-Means sont nombreux, il est notamment utilisé en :

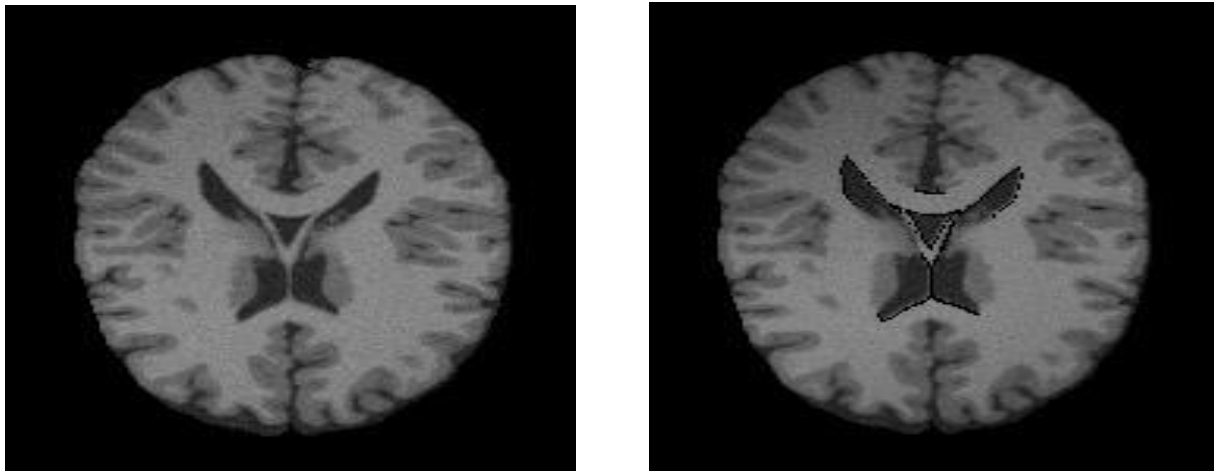
- la segmentation de la clientèle en fonction d'un certain critère (démographique, habitude d'achat etc....)
- Utilisation du clustering en Data Mining lors de l'exploration de données pour déceler des individus similaires. Généralement, une fois ces populations détectées, d'autres techniques peuvent être employées en fonction du besoin.
- Clustering de documents (regroupement de documents en fonction de leurs contenus. Pensez à comment Google Actualités regroupe des documents par thématiques)[41].

## 3-Principe de la méthode proposée :

Les méthodes classiques de segmentation basées sur l'apprentissage automatique, utilisent soit tout le volume IRM, soit toute une coupe (slice) de ce volume pour effectuer une classification ou un clustering des voxels ou pixels. Dans les deux cas, ceci pose un problème à cause d'un phénomène propre aux IRM et qui consiste au non uniformité de l'intensité dans les IRM. En effet, les pixels qui appartiennent au même tissu, mais qui sont éloignés dans l'image, peuvent avoir des intensités suffisamment différentes, pour que ces pixels soient affectés à des classes différentes (figure 31).

Pour remédier à ce problème, nous proposons d'effectuer un clustering local, en appliquant l'algorithme k-means à plusieurs sous-régions locales, puis fusionner les différents résultats locaux pour produire la segmentation finale de l'IRM.

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

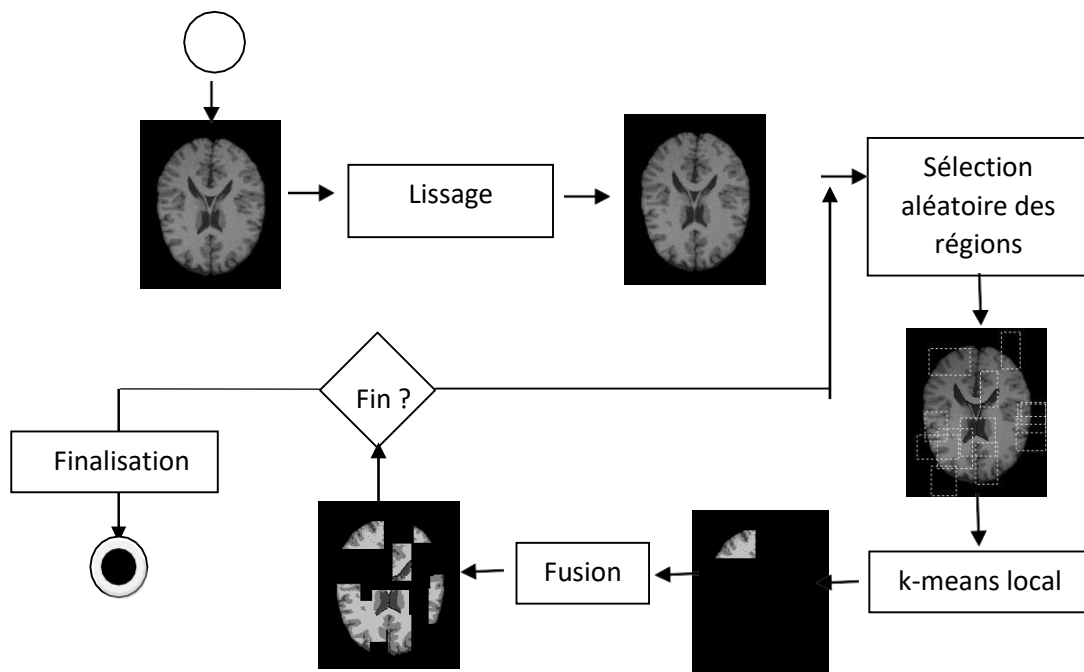


IRM avec uniformité d'intensité

IRM avec non uniformité d'intensité

**Figure III.6 : Non uniformité d'intensité dans les IRM**

L'organigramme suivant montre le déroulement de la méthode proposée.



**Figure III.7 : Déroulement de la méthode proposée**

# CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

## 4- Étapes de la méthode :

### 4-1- Filtrage de bruit :

Le filtrage de bruit se fait par lissage de la coupe IRM par la moyenne des intensités des pixels voisins. Le lissage est appliqué à toute l'image d'une manière adaptative dans le sens où la moyenne est calculée en considérant uniquement les pixels voisins et dont l'intensité est proche de celle du pixel en question. Ceci permet de préserver les contours des tissus dans l'IRM, et de réduire le flou dans l'image qui en résulte.

#### Algorithme Lissage\_adaptatif

**Début**

**Pour**  $x \leftarrow 1$  à Largeur IRM -2 **Faire**

**Pour**  $y \leftarrow 1$  à Hauteur IRM -2 **Faire**

Somme  $\leftarrow 0$  Nombre  $\leftarrow 0$

**Pour**  $x' \leftarrow x-1$  à  $x+1$  **Faire**

**Pour**  $y' \leftarrow y-1$  à  $y+1$  **Faire**

**Si**  $|IRM[x,y] - IRM[x',y']| < \text{Seuillissage}$  **Alors**

Somme  $\leftarrow$  Somme +  $IRM[x',y']$

Nombre  $\leftarrow$  Nombre + 1

**Fin Si**

**Fin Pour**

**Fin Pour**

$IRM'[x,y] \leftarrow$  Somme / Nombre

**Fin Pour**

**Fin Pour**

$IRM \leftarrow IRM'$

**Fin**

### 4-2-Sélection de régions aléatoires :

Une série de régions aléatoires sont sélectionnées, où dans chacune on applique l'algorithme k-means. À chaque itération de l'algorithme principal, une région est positionnée aléatoirement

## CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

une position  $(x_r, y_r)$ , et de dimension aussi aléatoire  $W_r$  (largeur) et  $H_r$  (hauteur).

- $W_r = 1 + \text{random}(\text{LargeurIRM} - 50)$
- $H_r = 1 + \text{random}(\text{HauteurIRM} - 50)$
- $X_r = \text{random}(\text{LargeurIRM} - W_r - 1)$
- $Y_r = \text{random}(\text{HauteurIRM} - H_r - 1)$

Les expressions précédentes permettent d'avoir des largeurs des régions  $W_r$  comprises entre 20 et la largeur de l'IRM, des hauteurs des régions  $H_r$  comprises entre 20 et la hauteur de l'IRM. Les positions  $(x_r, y_r)$  sont d'office dans l'IRM.

### 4-3- Application locale de k-means :

L'algorithme de clustering k-means est appliqué dans une région locale, positionnée en  $(x_r, y_r)$  et de dimension  $W_r \times H_r$ . Le nombre de classe est 4, représentant les tissus suivants :

- 0: Fond de l'IRM (noir)
- 1: Liquide céphalo-rachidien (très foncé)
- 2 : Matière grise (foncé)
- 3 : Matière blanche (moins foncé à claire)

### 4-4-Fusion des résultats :

Tout d'abord, il faut noter qu'un pixel donné de la coupe IRM est classifié plusieurs fois selon son appartenance à une région locale aléatoirement sélectionnée. Une matrice de fusion, appelée MF, contient en chaque point  $(x, y)$  un vecteur de 4 éléments entiers, indexés de 0 à 3, représentant les 4 tissus rencontrés dans l'IRM. En chaque position  $i$  de ce vecteur est cumulé le nombre d'occurrences ou le pixel  $(x, y)$  était classifiée comme appartenant au tissu  $i$  ( $i \in \{0, 1, 2, 3\}$ ).

En supposant que l'algorithme k-means a produit la classification des pixels de la région locale courante, et l'a stocké dans une matrice classes  $[. , .]$  de dimensions  $W_r \times H_r$ . L'algorithme suivant représente la fusion des résultats obtenus à cette itération avec les résultats précédents, cumulé dans la matrice de fusion MF :

## CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales

**Pour**  $x \leftarrow x_r$  à  $x_r+W_r-1$  **Faire**

**Pour**  $y \leftarrow y_r$  à  $y_r+H_r-1$

**Faire**

$MF[x,y,classe[x-x_r,y-y_r]] \leftarrow MF[x,y,classe[x-x_r,y-y_r]]+1$

**Fin Pour**

**Fin Pour.**

La boucle sur la génération de régions locales, de clustering k-means et de fusion est répétée un nombre élevé de fois, de sorte qu'on soit sûr que tout pixel de l'image a appartenu aumoins à une région (ou plus) des régions aléatoirement générées.

### 4-5-Finalisation :

Cette étape consiste à calculer la classe finale de chaque pixel de la coupe IRM, et ce en se basant sur la matrice de fusion MF. La classe finale affectée au pixel à la position (x,y) de la coupe IRM est le rang correspondant au maximum du vecteur de la matrice de fusion MF à la position (x,y).

#### Algorithme Finalisation

**Début**

**Pour**  $x \leftarrow 0$  à  $\text{LargeurIRM}-1$  **Faire**

**Pour**  $y \leftarrow 0$  à  $\text{HauteurIRM}-1$  **Faire**

Maximum  $\leftarrow MF[x,y,0]$

Rang  $\leftarrow 0$

**Pour**  $i \leftarrow 1$  à  $3$  **Faire**

**Si**  $MF[x,y,i] > \text{Maximum}$  **Alors**

Maximum  $\leftarrow MF[x,y,i]$

Rang  $\leftarrow i$

**Fin Si**

**Fin Pour**

ClasseFinale[x,y]  $\leftarrow$  Rang

**Fin Pour**

**Fin Pour**

## **CHAPITRE 3 : Moyennes Mobiles Locales pour la Segmentation d'IRM Cérébrales**

La matrice Classe Finale contient les classes finales de pixels.

### **5- Conclusion :**

Dans ce chapitre, nous avons présenté une nouvelle méthode d'utilisation de l'algorithme k-means pour la segmentation d'IRM cérébrales. La méthode est basée sur le clustering de pixels appartenant aux différentes régions locales, puis à la fusion des résultats obtenus. Au chapitre suivant, nous présenterons l'implémentation et l'expérimentation de la méthode proposée.

# CHAPITRE 04

## Implémentation Et Test

# CHAPITRE 4 : Implémentation et Test

## 1. Introduction :

Ce chapitre porte sur l'expérimentation d'application de segmentation d'image médicale (IRM de cerveau). Tout d'abord, avant de définir l'implémentation de l'algorithme K-means, nous présentons l'environnement matériel et la plate-forme logicielle utilisée pour ce travail et enfin, avant de passer au test en montrant les résultats d'expérimentations.

## 2. Le langage Python

### 2.1- A propos du langage Python

Python est un langage de programmation de haut niveau interprété pour la programmation à usage général. Créé par **Guido van Rossum**, et publié pour la première fois en 1991. Python repose sur une philosophie de conception qui met l'accent sur la lisibilité du code, notamment en utilisant des espaces significatifs. Il fournit des constructions permettant une programmation claire à petite et grande échelle [42].

Python propose un système de typage dynamique et une gestion automatique de la mémoire. Il prend en charge plusieurs paradigmes de programmation, notamment orienté objet, impératif, fonctionnel et procédural, et dispose d'une bibliothèque standard étendue et complète.

Python est un langage de programmation open-source et de haut niveau, développé pour une utilisation avec une large gamme de systèmes d'exploitation. Il est qualifié de langage de programmation le plus puissant en raison de sa nature dynamique et diversifiée. Python est facile à utiliser avec une syntaxe super simple très encourageante pour les apprenants débutants, et très motivante pour les utilisateurs chevronnés [42].

Il est le langage de programmation open source le plus couramment utilisé par les informaticiens. Ce langage se démarque dans les domaines de la gestion d'infrastructures, de l'analyse de données ou du développement de logiciels. En fait, dans ses qualités, Python permet aux développeurs de se concentrer sur ce qu'ils font, et non sur la façon dont ils le font. Il libère les développeurs des contraintes formelles des anciens langages qui tourmentaient leur époque. Par conséquent, développer du code en Python est plus rapide que dans d'autres langages.

Il est également disponible pour les débutants si nous passons du temps à nous lancer. De nombreux tutoriels sont également disponibles sur des sites dédiés. Les réponses aux questions peuvent toujours être trouvées sur les forums informatiques, car de nombreux professionnels l'utilisent.

# CHAPITRE 4 : Implémentation et Test

## 2.2 Quelles sont les principales raisons qui poussent à apprendre Python ? [42].

- Utilisé par des sites web pionniers : tels que Microsoft, YouTube, Drop Box,... Python a une forte demande sur le marché.
- Richesse en outils : de nombreux IDE sont dédiés au langage Python : Pycharm, Wing, PyScripter, Spyder...
- Python est orienté objet : la puissance du langage python est fortement marquée par son aspect orienté objet, qui permet la création et la réutilisation de codes. En raison de cette possibilité de réutilisation, le travail est effectué efficacement et réduit beaucoup de temps. Au cours des dernières années, la programmation orientée objet s'est rapporté non seulement à des classes et des objets, mais à de nombreuses bibliothèques et frameworks. Python a son tour connu dans ce contexte un grand essor : des dizaines de milliers de bibliothèques sont disponibles à l'aide de l'outil pip de gestion des packages.
- Simplicité et lisibilité du code : Python a une syntaxe simple qui le rend approprié pour apprendre la programmation en tant que premier langage. L'apprentissage est plus fluide et rapide que d'autres langages tels que Java, qui nécessite très tôt une connaissance de la programmation orientée objet ou du C/C++ qui nécessite de comprendre les pointeurs. Néanmoins, il est possible d'en apprendre davantage sur la programmation orientée objet en Python lorsqu'il est temps. Par conséquent, Python peut être utilisé comme prototype et peut être implémenté dans un autre langage de programmation après avoir testé le code.
- Python est open source donc gratuit : Python étant un langage de programmation open source, il est gratuit et permet une utilisation illimitée. Avec cette licence open source, il peut être modifié, redistribué et utilisé commercialement. Avec cette licence, Python est devenu robuste, doté de capacités évolutives et portables et est devenu un langage de programmation largement utilisé.
- Python est multiplateforme : Python peut être exécuté sur tous les principaux systèmes d'exploitations, tels que : Mac OS, Microsoft Windows, Linux et Unix... Ce langage de programmation offre une meilleure expérience de travail avec n'importe quel système d'exploitation.
- Python est très puissant en terme de production : la puissance du langage Python a été démontré sur le terrain du développement :
  - Développement Web, en utilisant les frameworks Django, Flask, Pylons
  - Science des données et visualisation à l'aide de Numpy, Pandas et Matplotlib
  - Applications de bureau avec Tkinter, PyQt, Gtk, wxWidgets et bien d'autres..
  - Applications mobiles utilisant Kivy ou BeeWare

## CHAPITRE 4 : Implémentation et Test

– Education : Python est un excellent langage pour apprendre l’algorithmique et la programmation ! Par conséquent largement utilisé aux Lycées, Classes préparatoires, Instituts supérieurs, Universités...

### 2.3 Bibliothèques en Python :

Normalement, une bibliothèque est une collection de livres ou une pièce ou un lieu où de nombreux livres sont stockés pour être utilisés plus tard. De même, dans le monde de la programmation, une bibliothèque est une collection de codes précompilés qui peuvent être utilisés ultérieurement dans un programme pour certaines opérations spécifiques bien définies. Outre les codes précompilés, une bibliothèque peut contenir de la documentation, des données de configuration, des modèles de message, des classes et des valeurs, etc. [43]

Une bibliothèque Python est une collection de modules associés. Il contient des paquets de code qui peuvent être utilisés à plusieurs reprises dans différents programmes. Cela rend la programmation Python plus simple et pratique pour le programmeur. Comme nous n'avons pas besoin d'écrire le même code encore et encore pour différents programmes. Les bibliothèques Python jouent un rôle très important dans les domaines de l'apprentissage automatique, de la science des données, de la visualisation des données, etc. [43]

### 2.4 Principales Bibliothèques utilisés :

**Matplotlib** : cette bibliothèque est chargée de tracer des données numériques. Et c'est pourquoi il est utilisé dans l'analyse des données. C'est également une bibliothèque open-source et trace des figures haute définition comme des camemberts, des histogrammes, des nuages de points, des graphiques, etc.

**Pandas** : Les pandas sont une librairie importante pour les data scientists. Il s'agit d'une bibliothèque d'apprentissage automatique open source qui fournit des structures de données flexibles de haut niveau et une variété d'outils d'analyse. Il facilite l'analyse des données, la manipulation des données et le nettoyage des données. Les pandas prennent en charge des opérations telles que le tri, la réindexation, l'itération, la concaténation, la conversion de données, les visualisations, les agrégations, etc.

**Numpy** : Le nom « Numpy » signifie « Numerical Python ». C'est la bibliothèque couramment utilisée. Il s'agit d'une bibliothèque d'apprentissage automatique populaire qui prend en charge de grandes matrices et des données multidimensionnelles. Il se compose de fonctions mathématiques intégrées pour des calculs faciles. Même des bibliothèques comme TensorFlow utilisent Numpy en interne pour effectuer plusieurs opérations sur les tenseurs. Array Interface est l'une des principales caractéristiques de cette bibliothèque.

# CHAPITRE 4 : Implémentation et Test

## 3. Environnement de développement :

### 3.1-Matériel utilisé :


Notre travail s'est effectué sur Laptop DELL ayant les caractéristiques suivantes :

- Processeur : Intel(R) Core (TM) i5-6300U CPU @ 2.40GHz 2.50 GHz
- ID de périphérique : D4DA02F9-543F-4680-83D6-393C6C7CEF67
- ID de produit : 00330-50311-82708-AAOEM
- Mémoire RAM installée : 8,00 Go
- Ecran : 14 pouces.
- Type de Système : Windows10 , 64 bits, Processeur x64.


### 3.2-Outils de développement :

Nous avons travaillé dans ce chapitre pour créer un système de filtrage dans lequel l'algorithme est codé sera en ligne, en utilisant Google Colab (abréviation de Colaboratory) avec langage de programmation python.

#### 3.2.1 -Google Colab :

 Colab (abréviation de Colaboratory) est un produit proposé par Google Research qui permet chercheurs en apprentissage automatique pour travailler sur des projets dans le navigateur. Similaire à Google Docs, Il nous permet de partager des projets entre plusieurs personnes, et le meilleur de tous, qu'il offre un accès gratuit aux CPU afin que nous puissions former des modèles rapidement sans aucun enregistrement. [44].

Comme définition générale, Colab est un environnement de notebook gratuit qui s'exécute entièrement dans le cloud. Mieux encore, il ne nécessite aucune configuration et les blocs- notes que nous créons peuvent être modifiés simultanément par les membres de notre équipe, tout comme nous modifions des documents dans Google Docs. Colab prend en charge de nombreuses bibliothèques de machine learning populaires qui peut être facilement chargées sur notre ordinateur.

 Colab est un produit de type Jupyter Notebook de Google Research. Un développeur de programme Python peut utiliser ce bloc-notes pour écrire et exécuter des codes de programme Python aléatoires simplement à l'aide d'un navigateur Web [45].

En un mot, Colab est une version hébergée dans le cloud de Jupyter Notebook. Pour utiliser Colab, vous n'avez pas besoin d'installer et d'exécuter ou de mettre à niveau votre matériel informatique pour répondre aux exigences de charge de travail intensives du CPU/GPU de Python. De plus, Colab vous donne un accès gratuit à l'infrastructure informatique comme le stockage, la

## CHAPITRE 4 : Implémentation et Test

mémoire, la capacité de traitement, les unités de traitement graphique (GPU) et les unités de traitement de tenseur (TPU).

Google a spécialement programmé cet outil de codage Python basé sur le cloud en gardant à l'esprit les besoins des programmeurs d'apprentissage automatique, des analystes de données volumineuses, des scientifiques des données, des chercheurs en IA et des apprenants Python [45].

La meilleure partie est un bloc-notes de code pour tous les composants nécessaires pour présenter un apprentissage automatique complet ou science des données projet aux superviseurs ou aux parrains du programme. Par exemple, votre bloc-notes Colab peut contenir des codes exécutables, des codes Python en direct, du texte enrichi, HTML, LaTeX, des images, des visualisations de données, des graphiques, des tableaux, etc.

### 3.2.2-Les offres de Google Colab :

En tant que programmeurs, nous pouvons effectuer les opérations suivantes à l'aide de Google Colab:

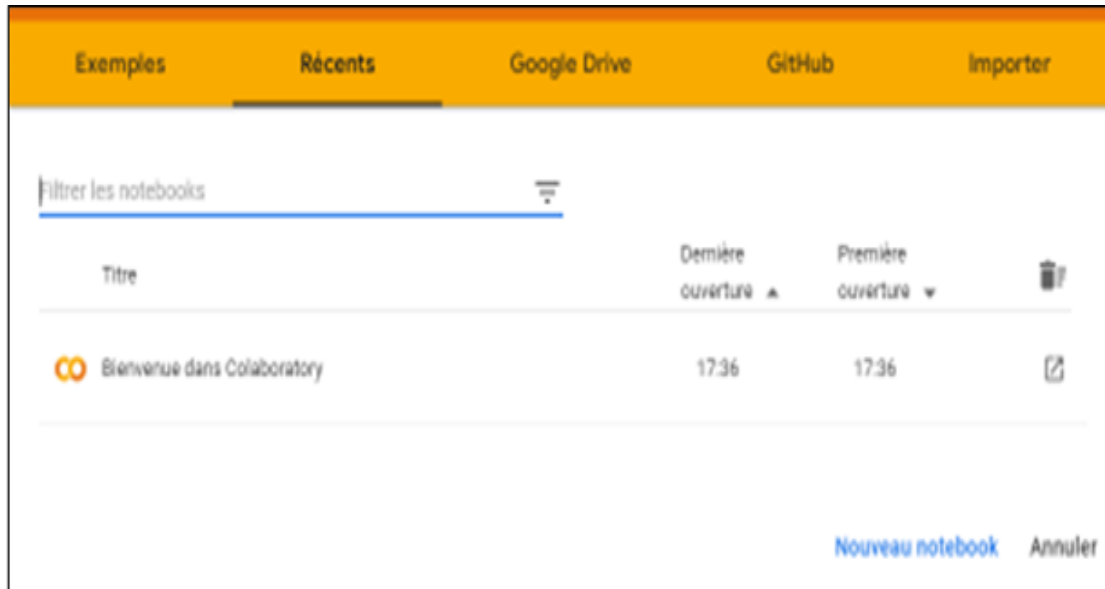
- Écrire et exécuter du code en Python.
- Documenter notre code qui prend en charge les équations mathématiques.
- Créer/Télécharger/Partager des blocs-notes.
- Importer/Enregistrer des blocs-notes depuis/vers Google Drive.
- Importer/publier des notebooks depuis GitHub.
- Importer des ensembles de données externes.
- Intégrer PyTorch, TensorFlow, Keras, OpenCV.
- Service Cloud gratuit avec GPU gratuit.

### 4. Création de projet et l'explication de l'interface :

Pour créer un nouveau projet dans Google Colab, nous ouvrons l'URL suivante dans notre navigateur : <https://colab.research.google.com>.

Notre navigateur affichera l'écran suivant (En supposant que nous sommes connectés à notre Google Drive) :

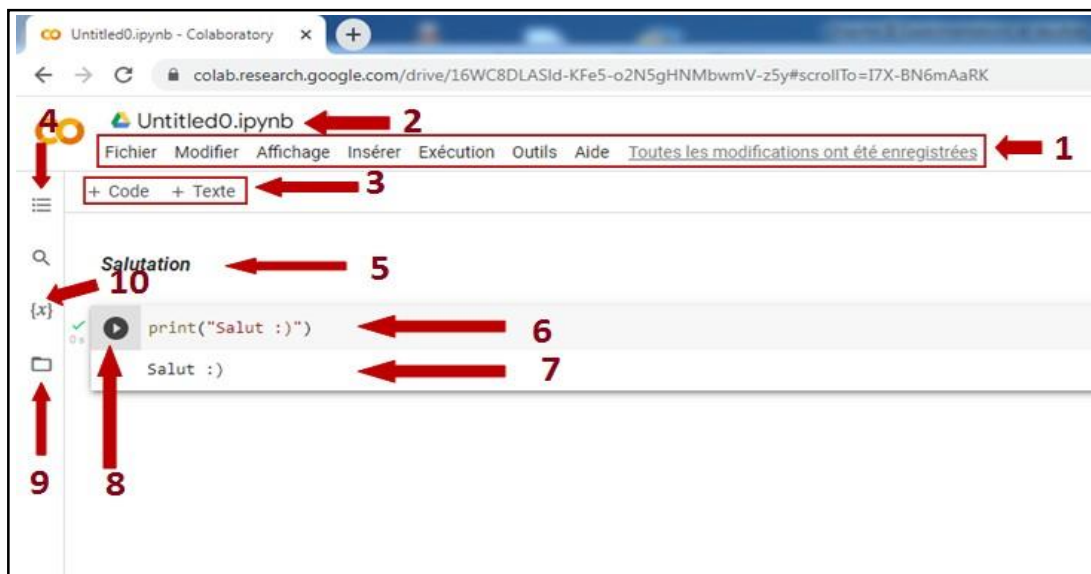
## CHAPITRE 4 : Implémentation et Test



**Figure IV.01 : Création un projet dans Google Colab**

**Remarque :** Nous devons avoir un compte Gmail pour utiliser cet outil. Si nous n'avons pas encore de compte Gmail, nous devons en créer un. Nous pouvons créer un nouveau fichier en cliquant sur NOUVEAU NOTEBOOK. Ensuite, l'interface de travail apparaîtra.

❖ L'explication de l'interface



**Figure IV.02 : Partie gauche de l'interface**

**1. Barre de menus :** comme dans toute autre application, cette barre de menus peut être utilisée pour manipuler le fichier entier ou ajouter de nouveaux fichiers. Permet de parcourir les différents onglets et de se familiariser avec les différentes options.

**2. Nom de fichier :** il s'agit du nom de votre fichier. Il peut être cliqué pour changer le nom. Il est

## CHAPITRE 4 : Implémentation et Test

préférable de ne pas faire le changement de l'extension (.ipynb) lors de la modification du nom car cela peut rendre le fichier qui ne s'ouvre pas.

**3. Insérer une cellule de code ou texte :** Il se compose de deux boutons à travers lesquels nous pouvons ajouter un code ou une cellule de texte sous la cellule précédente ou spécifique.

**4. Table des matières :** Ici, nous serons en mesure de créer et de parcourir différentes sections à l'intérieur de notre bloc-notes. Les sections nous permettent d'organiser les instructions logicielles et d'améliorer la possibilité de lire.

**5. Cellule de texte :** C'est ici que nous pouvons écrire un texte ou un titre expliquant la cellule de code.

**6. Cellule de code :** C'est ici que nous pouvons écrire notre code ou algorithme utilisé.

**7. Sortie :** il s'agit de la sortie de votre code, y compris les erreurs, qui seront affichées.

**8. Exécuter la cellule :** il s'agit du bouton d'exécution. Cliquer dessus exécutera tout code inséré dans la cellule à côté.

**Remarque :** Il y a deux indications lors de l'exécution du code



: Marque verte pour le correct processus d'exécution de code.

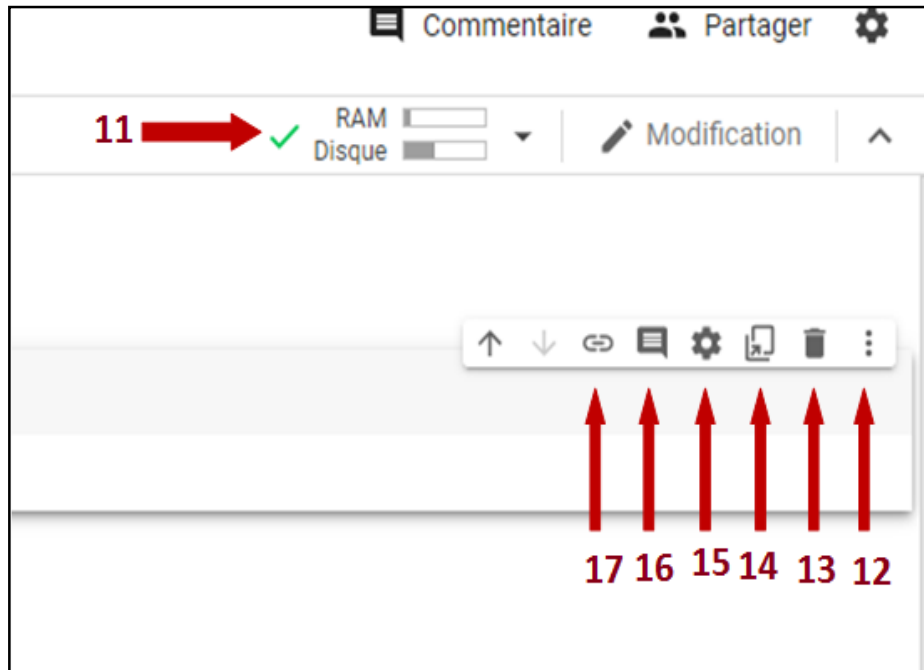


: Marque rouge pour le mauvais processus d'exécution de code.

**9. Fichiers :** ici, nous serons téléchargés des ensembles de données et d'autres fichiers depuis notre ordinateur et Google Drive.

**10. Extraits de code :** ici, nous pourrons trouver des extraits de code pré-écrits pour différentes fonctionnalités, telles que l'ajout de nouvelles bibliothèques ou le référencement d'une cellule à une autre.

## CHAPITRE 4 : Implémentation et Test



**Figure IV.03 : Partie droite de l'interface**

- 11. RAM et disque :** Tout le code que nous écrivons sera exécuté sur un ordinateur de Google et nous ne verrons que la sortie. Cela signifie que même si nous avons un ordinateur lent, le fonctionnement de grandes parties du code n'est pas un problème. Google spécialise une certaine quantité d'espace RAM et disque pour chaque utilisateur.
- 12. Plus d'options :** Contient des options pour couper et copier une cellule ainsi que l'option pour ajouter un formulaire et masquer le code.
- 13. Supprimer la cellule :** ce bouton supprimera la cellule sélectionnée.
- 14. Miroir la cellule :** mettre en miroir la cellule dans un onglet.
- 15. Paramètres :** ce bouton nous permettra de modifier le thème du notebook, le type de police, la taille, la largeur de l'indentation, etc.
- 16. Commentaire :** Ce bouton nous permettra de créer un commentaire sur la cellule spécifiée. Ce sera un commentaire sur la cellule et non un commentaire dans la cellule.
- 17. Lien vers la cellule :** ce bouton créera une URL qui sera liée à la cellule que nous avons sélectionnée.

# CHAPITRE 4 : Implémentation et Test

## 4-1-Extraits du code :

Modules utilisés

```
from PIL import Image
from matplotlib import pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.cluster import KMeans
from sklearn.metrics import classification_report, confusion_matrix
import pandas as pd
from random import*
```

## 4-2-Chargement, filtrage et affichage IRM :

```
❖ #chargement
❖ Largeur = 181
❖ Hauteur = 217
❖ nomfichier="/content/drive/MyDrive/Slice83.rawb"
❖ f=open(nomfichier,"rb")
❖ tableau = f.read()
❖ image = bytearray(tableau)
❖ f.close()
❖
❖ #filtrage
❖ def Im2d(x,y):
❖     return image[y*Largeur+x]
❖ for y in range(1,Hauteur-1):
❖     for x in range(1,Largeur-1):
❖         somme=0
❖         for yy in range(y-1,y+2):
❖             for xx in range(x-1,x+2):
❖                 somme=somme+Im2d(xx,yy)
❖         image[y*Largeur+x]=somme//9
❖
❖ #affichage
❖ print("Image brute")
❖ imaff=Image.new('RGB', (181,217))
❖ for y in range(Hauteur):
❖     for x in range(Largeur):
❖         intensite= image[y*Largeur+x]
❖         imaff.putpixel((x,y), (intensite,intensite,intensite))
❖ plt.figure()
❖ plt.imshow(imaff)
```

## CHAPITRE 4 : Implémentation et Test

### 4-3-Chargement données d'apprentissage :

```
❖ nomfichier="/content/drive/MyDrive/rt83.rawb"  
❖ f=open(nomfichier,"rb")  
❖ tableau = f.read()  
❖ rt = bytearray(tableau)  
❖ f.close()
```

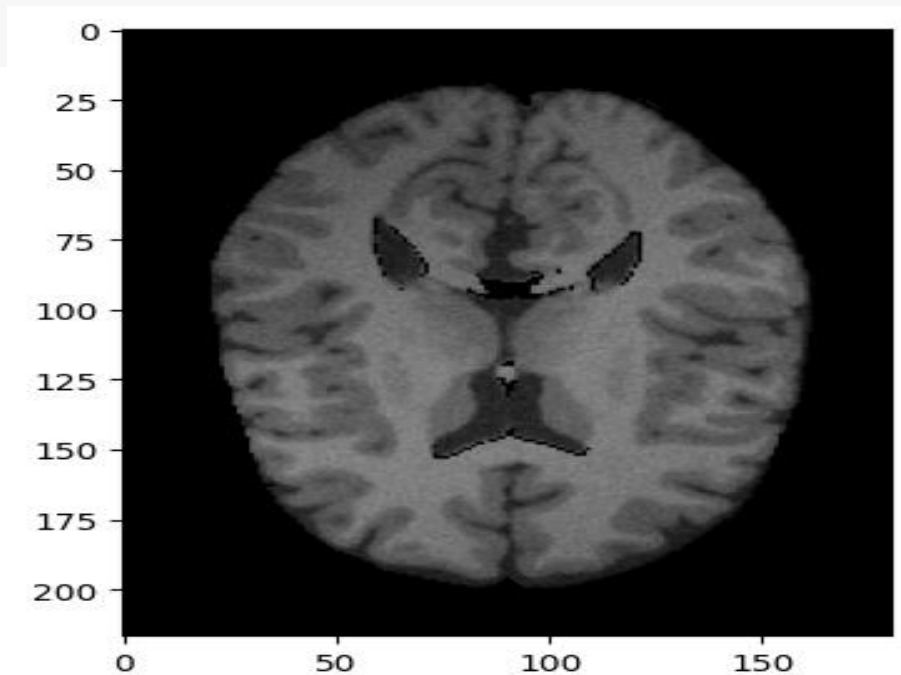


Figure IV.04 :Image brute à segmenter

### 4-4-Initialisation des données d'apprentissage et sélection des sous images :

```
❖ XX=np.array(image)  
❖ YY=np.array(rt)  
❖ #matrice global des étiquettes  
❖ ZZ=np.zeros(Largeur*Hauteur)  
❖  
❖ #début de bouclage sur kmeans  
❖ for iteration in range(10):  
❖     xd=randint(1,50)  
❖     xf=randint(Largeur-50, Largeur-1)  
❖     yd=randint(0,50)  
❖     yf=randint(Largeur-50, Largeur-1)
```

Dans cet extrait de code, on montre comment l'algorithme du k-means est exécuté 10 fois, chaque fois dans une région de l'image définie par les points (xd,yd) et (xf,yf).

## CHAPITRE 4 : Implémentation et Test

### 4-5-K-meand local et copie des résultats :

```
❖ Largeur2 = xf-xd+1
❖ Hauteur2 = yf-yd+1
❖
❖ km = KMeans(n_clusters=4)
❖ km.fit(X.reshape(-1,1))
❖
❖ Y_predict = km.labels_
❖ print(confusion_matrix(Y,Y_predict))
❖ print(classification_report(Y,Y_predict))
❖ print(Y_predict.size)
❖
❖ # copie des labels sur la matrice ZZ
❖ for y in range(Hauteur2):
❖     for x in range(Largeur2):
❖         ZZ[Largeur*(y+yd)+x+xd]= Y_predict[y*Largeur2+x]
❖
```

### 4-6-Affichage du résultat global :

```
❖ #Affichage résultat global
❖ print("Image gobale segmentée par kmeans")
❖ imkmeansglob=Image.new('RGB', (Largeur,Hauteur))
❖ for y in range(Hauteur):
❖     for x in range(Largeur):
❖         intensite= int(30*ZZ[y*Largeur+x])
❖         imkmeansglob.putpixel((x,y), (intensite,intensite,intensite))
❖ plt.figure()
❖ plt.imshow(imkmeansglob)
❖
```

Cet extrait du code montre l'exécution de k-means sur la zone sélectionné et la fusion des résultats dans la matrice ZZ.

### 4-7-Exemple de matrice de confusion :

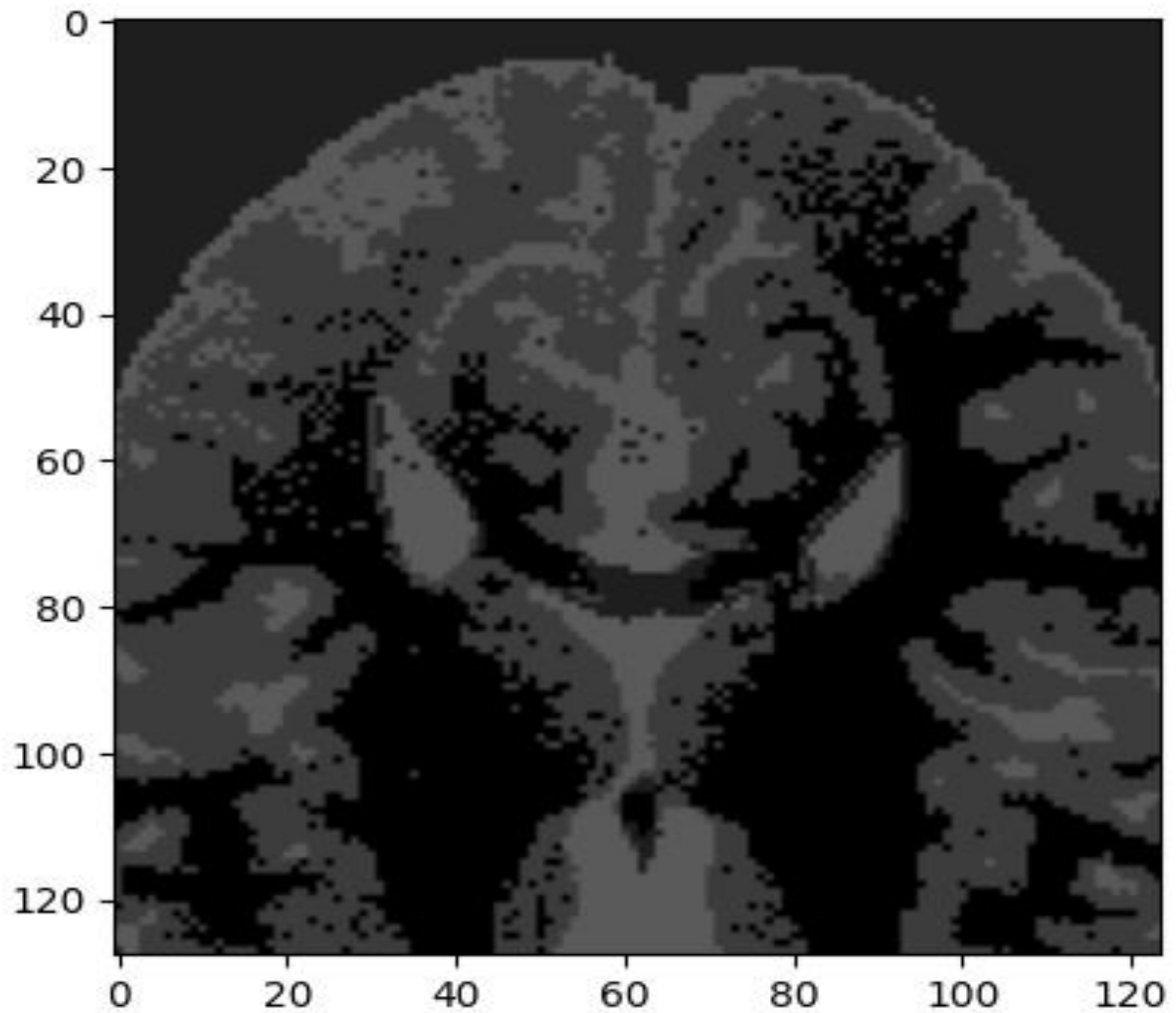
```
[[ 0 19407 0 0]
 [ 0 156 239 2476]
 [ 820 0 7285 932]
 [ 6211 0 1742 9]]
```

## CHAPITRE 4 : Implémentation et Test

### 4-8-Exemple de Métriques d'évaluation :

	precision	recall	f1-score	support
0	0.00	0.00	0.00	19407
1	0.01	0.05	0.01	2871
2	0.79	0.81	0.80	9037
3	0.00	0.00	0.00	7962
accuracy			0.19	39277
macro avg	0.20	0.22	0.20	39277
weighted avg	0.18	0.19	0.18	39277

### 4-9-Exemples de résultats de k-means locaux :



## CHAPITRE 4 : Implémentation et Test

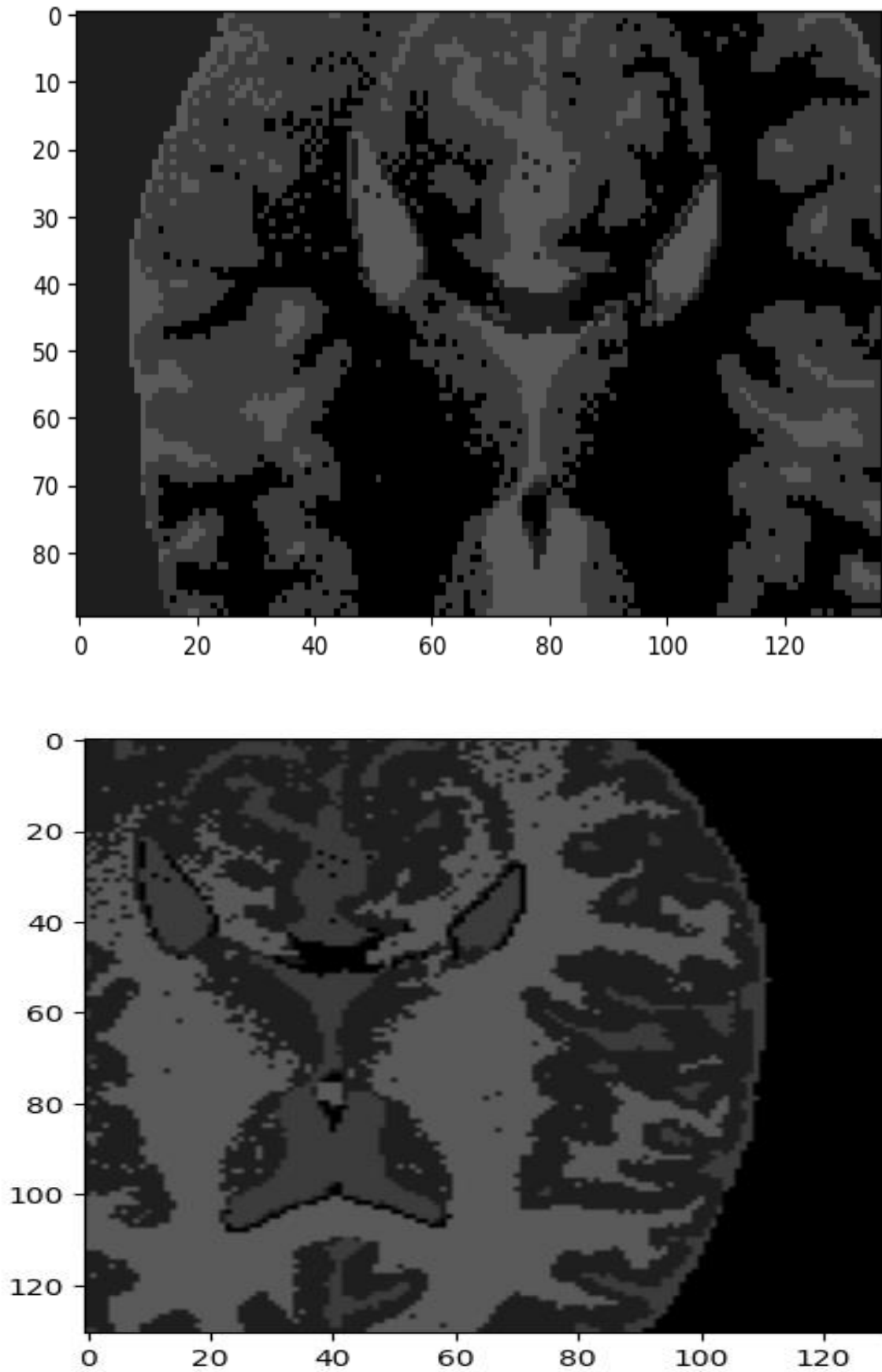
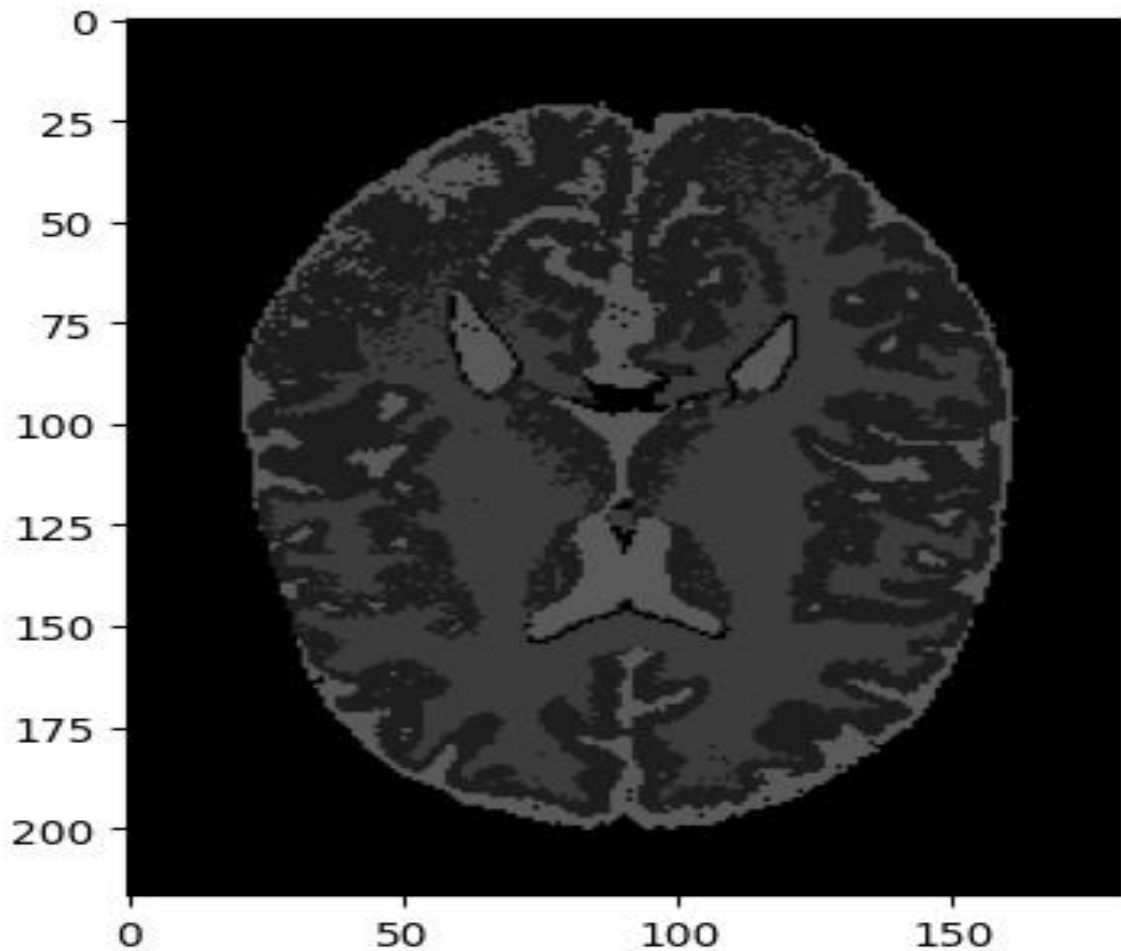


Figure IV.05 :03 Exemple des k-means locaux

## CHAPITRE 4 : Implémentation et Test

### 4-10-Résultat global :

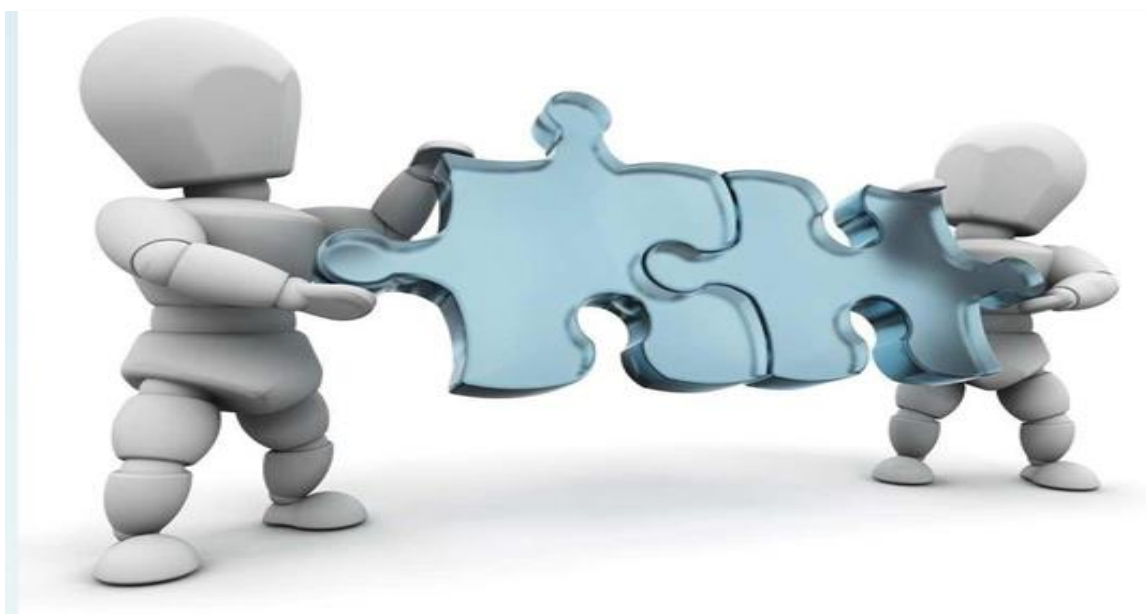


**Figure IV.06 : Image globale segmentée par k-means**

### 5-Conclusion :

Dans ce chapitre, nous avons montré les éléments essentiels de l'implémentation de notre méthode, ainsi qu'un exemple des résultats obtenus. Nous avons présenté également l'environnement de développement, à savoir Google Colab, qui est un environnement en ligne, facile à utiliser et ne nécessite aucune installation. Nous avons également présenté le langage Python et les principales bibliothèques utilisées. Les résultats obtenus montrent que la méthode proposée, basée sur l'exécution de k-means sur des régions locales de l'image, permet d'améliorer les résultats de cet algorithme.

# Conclusion Général



# Conclusion Générale

La segmentation d'images médicales en générale, et les IRM en particulier demeure problématique car ces images sont largement entachées par le bruit et par différentes déformation, notamment la non-uniformité des intensités, qui fausse les résultat de segmentation.

Afin de contribuer l'amélioration de la segmentation des IRM cérébrales, nous avons présenté dans ce mémoire de master une nouvelle méthode, basée sur l'application d'un clustering locale des données IRM, puis fusionner les résultats partiels, en obtenant ainsi un clustering global meilleur que celui obtenu par un clustering de la totalité de l'image

Nous avons choisi l'algorithme de clustering k-means, car il a été largement utilisé pour la segmentation des IRM. Le choix des régions locales s'effectue d'une manière aléatoire de sorte que les régions sélectionnées soient de petites tailles. Ceci permet de surmonter le problème de la non-uniformité des intensités dans les IRM. En effet, dans une région locale ce problème pourra être négligé, car, il n'affecte pas significativement les moyennes des classes.

Nous avons implémenté notre méthode en utilisant le langage Python sous l'environnement de développement Google Colab. Les premiers résultats visuels montrent le potentiel de la méthode proposée pour bien segmenter les IRM, et surtout en surpassant le problème de la non-homogénéité d'intensités dans les IRM.

Comme perspectives à ce travail, il est question de quantifier la performance de la méthode proposée, et de la comparer aux autres méthodes. Il est également intéressant d'utiliser d'autres algorithmes de clustering, tels que le DBSCAN.

# *Bibliographie*

## Bibliographie:

- [1] : M. Kass, A. Witkin, et D. Terzopoulos, "Snakes: Active contour models", International Journal of Computer Vision 1(4):321-331,1987.
- [2] : Généralités en anatomie pour les STAPS : Par Monique R 26 08 2018 (<https://www.elsevier.com/fr-fr/connect/anatomie/generalites-en-anatomie-pour-les-staps>)
- [3] : Germond, L. Trois principes de coopération pour la segmentation en imagerie de résonance magnétique cérébrale. Université Joseph FOURIER Grenoble1, 1999
- [4] : <https://digitalinsiders.feelandclic.com/construire/definition-quest-machine-learning/>
- [5] : Marref Nadia-thème : apprentissage incrémental et machines à vecteurs supports-université :Hadj lakhdar-batna-2013
- [6] : SèmiyouAyélé OSSENI -Doctorat de l'université de toulouse, Nanoplateformes Hybrides Multimodales Pour L'imagerie Médicale,2012
- [7] : [https://fr.wikipedia.org/wiki/Imagerie\\_m%C3%A9dicale](https://fr.wikipedia.org/wiki/Imagerie_m%C3%A9dicale).
- [8] : Belarouci Sara & Benmokhtar Sara- Master En Génie Biomédical, Thème Méthode Coopérative Pour La Segmentation D'images IRM Cérébrales Basée Sur Les Techniques FCM Et Level Set- Université Abou Bekr Belkaid Tlemcen -2011/2012
- [9] : Boussedi Rym Sabrine, Kenadil Meryem, « Segmentation des tumeurs cérébrales par la ligne de partage des eaux sur des images IRM », Master en Électronique Biomédicale, 2010-2011.
- [10] : <http://qcouturier.free.fr/TPESite/page/IRM.html>
- [11] : K. Aounallah, les approches de segmentation d'image par coopération régions contours, 2010.
- [12] : C.Houassine, segmentation d'images par une approche biomimétique hybride. Université m'hamed bougara- boumerdes. 2012.
- [13] : M. Kass, A. Witkin And D. Terzopoulos, snakes : active contour models, international journal of computer vision, 1(4): pp321–331, 1987. cc
- [14] : M. Melliani, segmentation d'image par coopération régions-contours, magistère en informatique, école nationale supérieure d'informatique, 2012.
- [15] : Sarra Benfriha et Asma Hamel-Segmentation d'image par Coopération région contours-Mémoire Master Professionnel Université Kasdi Merbah-Ouargla-2015/2106.
- [16] : V. Barra, and J. H. Boire, "Segmentation floue des tissus cérébraux en IRM 3D: une approche possibiliste versus autres méthodes," Rencontres Francophones sur la logique floue et ses applications, Valenciennes, Editions Cépaduès, pp. 193-198. 1999.
- [17] : Yuan DONG –Thèse: Modélisation probabiliste de classificateurs d'ensemble pour des problèmes à deux classes- Université De Technologie De Troyes-2013 ,

[18]:<https://jcrisch.wordpress.com/2015/05/04/valider-un-modele-statistique-avec-lacross-validation/>

[19] : Quentin Duché- Etude des effets de volume partiel en IRM cérébrale pour l'estimation d'épaisseur corticale-Université de Rennes1 -Thèse soutenue 2015

[20] : AISSAOUI Sihem -Université d'Oran Es-senia -Apprentissage automatique et sécurité des systèmes d'information Application : Un système de détection d'intrusion basé sur les Séparateurs à Vaste Marge (SVM)-2007/2008.

[21] : M.Melliani, segmentation d'image par coopération régions-contours, magistère en informatique, école national supérieur d'informatique, 2012.

[22] : MAITRE, H. (2004). Le traitement des images.

[23] : Gilbert Saporta, « Data Mining » ou fouille de données , RST « Epidémiologie »,4.2 Data Mining 5/12/04.

[24] : Stéphane Tuffery- Cours De Data Mining Université Rennes 1 Master 2 Ingénierie économique et financière 7 février 2014

[25] : Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From data mining to knowledge discovery in databases," Dans *mag KDD overview*, pp 1–34, 1996.

[26]: Boucheham Bachir – cours 2eme année master Académique 'data Manning' –Université de skikda 20 aout 55 -2016

[27] : LAMICHE Chaabane- thèse de doctorat En Sciences Spécialité : Informatique Thème "fusion et fouille de données guidées par les connaissances : application a l'analyse d'image" 18/06/2013

[28] : Pdf l'intelligence artificielle au où en sommes-nous? canada -conseil des technologies de l'information et des communications avril 2015.

[29] : C.Houassine, segmentation d'images par une approche biomimétique hybride. Université m'hamed bougara- boumerdes. 2012.

[30] : Belarouci Sara & Benmokhtar Sara- Master En Génie Biomédical, Thème Méthode Coopérative Pour La Segmentation D'images IRM Cérébrales Basée Sur Les Techniques FCM Et Level Set- Université Abou Bekr Belkaid Tlemcen -2011/2012

[31] : MEZIANE Abdelfettah \_thème de magister: Etude et comparaison des méthodes de segmentation d'images cérébrales. Université Abou-Bakr Belkaid-Tlemcen - 2011

[32] : Dr. Abdelhamid DJEFFAL - Cours Fouille de données avancée-Master 2 IDM-Université Mohamed Khider – Biskra-Département d'Informatique -Année Universitaire 2014/2015.

[33] : P.K. Saha, J.K. Udupa. – Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation. *Computer Vision and Image Understanding*, vol. 77, 2000, pp. 145–174. P.K. Saha,

J.K. Udupa. – Fuzzy connected object delineation: Axiomatic path strength.

[34] : definition and the case of multiple seeds. Computer Vision and Image Understanding, vol. 83, 2001, pp. 275–295.

[35] : Salima Nebti, Université Ferhat Abass-Sétif, Département d'Informatique, thèse de Doctorat thème Reconnaissance de caractères Manuscrits par Intelligence Collective, Mars 2013.

[36] : D'Hondt Frédéric, El Khayati Brahim, Etude de méthodes de Clustering pour la segmentation d'images en couleurs. Faculté Poly technique de Mons, 5ème Electricité, Certificat Applicatifs Multimédia.

[37] : L. Houria, B. Oumessad, « Classification Des Feuilles de Plantes à Base de Moment de Hu », Mémoire Master Académique en informatique, 2014.

[38] : L. Jérémy B.SC. « Forage non supervisé de données pour la prédiction d'activité dans les habitats intelligentes ». L'université du Québec à Chicoutimi comme exigence partielle de la maîtrise en informatique, juin2013.

[39] : Z.Guellil et L.Zaoui,s Proposition d'une solution au problème d'initialisation cas du K-means. Université des sciences et de la technologie d'Oran MB, Université Mohamed Boudiaf USTO -BP 1505 El Mnaouer -ORAN – Algérie.

[40] : K. Mokrani, « Segmentation d'images par classifieurs non supervisés : Application à l'Imagerie par Résonance Magnétique (IRM) », Mémoire de Magister, Université A.Mira de Bejaia 2008.

[41] : B.youns, Tout ce que vous voulez savoir sur l'algorithme K-Means. 10 avril 2018.

[42] : Younes. Derfoufi Enseignant au CRMEF OUJDA, « Programmation en langage Python» 21 juin 2019

[43] : Geeksforgeeks : <https://www.geeksforgeeks.org/libraries-in-python/>,consulter le : 04/04/2022

[44] : <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-définition-et-utilisation-de-ce-langage-informatique/>consulter le : 04/04/2022

[45] : <https://geekflare.com/fr/google-colab/>