

**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**

**Spécialité : Systèmes Informatiques (SI)**

**Mémoire de Fin d'Etude En Vue de l'Obtention du Diplôme**

**De Master en Informatique**



**Thème**

**Techniques d'apprentissage Automatique  
pour la segmentation d'images médicales  
(IRM)**

**Réaliser par :**

**Encadrer par : Professeur.Mazouzi Smaïne**

- ❖ Slimane Tich Tich Habiba
- ❖ Mosbah Ines

**Année Universitaire 2023/2024**

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



# *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*

***M. Mazouzi Smaine***

*Nous tenons, tout particulièrement et très sincèrement, à lui remercier de nous avoir proposé le sujet et de nous avoir encadré, 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 aux membres de Jury pour l'honneur qu'ils nous font d'avoir bien voulu lire ce Travail et de le juger.*

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

*Nous souhaitons 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 mes chères sœurs et Frères chacun par son nom.*

*À mes respectés chefs de service :*

*« **Djizi Boudjema** » et « **Kadoum Nouredine** »*

*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 chacun par son nom au :*

*« **Sous-direction des moyen généraux** »*

*Avec les quelles 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***



**HABIBA**

# *Dédicace*

*À mes très chers parents*

*Pour tout leur amour et toute leur tendresse*

*Je prie ALLAH pour les protéger,*

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

*À mes chères filles : **Chahd Djamila & Allaa Errahmene***

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

*Partager ma joie*



*Inès*

## Abstract :

Image segmentation, particularly for medical images, remains problematic due to noise and various distortions. In this master's thesis, we propose a new method for segmenting brain MRIs by pixel classification, using a non supervised approach with the k-means algorithm. This method is based on classifying pixels within local regions, where the issue of Intensity Non- Uniformity (INU) can be ignored. However, the partial results must be merged using a majority voting technique.

## Résumé :

La segmentation des images, en particulier les images médicales, reste problématique en raison du bruit et des diverses déformations. Dans ce mémoire de master, nous proposons une nouvelle méthode de segmentation des IRM cérébrales par classification des pixels, en utilisant une approche non supervisée avec l'algorithme des k-means. Cette méthode repose sur la classification des pixels dans des régions locales, où le problème de la non-uniformité des intensités (INU : Intensity Non Uniformity) peut être négligé. Toutefois, les résultats partiels doivent être fusionnés en utilisant une technique de vote majoritaire.

## ملخص:

تقسيم الصور، وخاصة الصور الطبية، لا يزال يمثل مشكلة بسبب الضوضاء والتشوهات المختلفة. في هذه الرسالة للماستر، نقترح طريقة جديدة لتقسيم صور الرنين المغناطيسي للدماغ من خلال تصنيف البكسلات، باستخدام نهج غير مُشرف مع خوارزمية k-means. تعتمد هذه الطريقة على تصنيف البكسلات في مناطق محلية، حيث يمكن تجاهل مشكلة عدم انتظام الكثافة (INU: Intensity Non Uniformity) ومع ذلك، يجب دمج النتائج الجزئية باستخدام تقنية التصويت بالأغلبية.

# Sommaire

| Désignation  | N° Page   |
|--|-----------|
| <b>Abstract</b>                                    |           |
| <b>Résumé</b>                                      |           |
| <b>ملخص</b>  |           |
| <b>Sommaire</b>                                    |           |
| <b>Table de figure</b>                             |           |
| <b>Introduction générale</b>                       |           |
| <b>Chapitre 1 : Segmentation D'Images Médicale</b> |           |
| 1-Introduction                                     | <b>3</b>  |
| <b>Partie 1 : Imagerie médicales</b>               | <b>3</b>  |
| 2-Notion générale sur les images                   | <b>3</b>  |
| 2-1 Définition d'image                             | <b>3</b>  |
| a. Image analogique                                | <b>4</b>  |
| b. Image numérique                                 | <b>4</b>  |
| 2-2 Types d'images                                 | <b>5</b>  |
| 1. Image matricielle (bitmap)                      | <b>5</b>  |
| 2. Image vectorielle                               | <b>5</b>  |
| Caractéristique d'une image numérique              | <b>6</b>  |
| 1. Pixel   | <b>6</b>  |
| 2. Bruit   | <b>7</b>  |
| 3. Dimension                                       | <b>7</b>  |
| 4. Le contraste                                    | <b>7</b>  |
| 5. Résolution                                      | <b>8</b>  |
| 6. Extension                                       | <b>8</b>  |
| 7. La luminance                                    | <b>8</b>  |
| 3. L'imagerie Médicale                             | <b>9</b>  |
| 3-1-Techniques de l'imagerie médicale              | <b>9</b>  |
| 3-1-1 Scanographie                                 | <b>9</b>  |
| 3-1-2 Scintigraphies (Médecine Nucléaire)          | <b>10</b> |
| 3-1-3 Echographie                                  | <b>10</b> |
| 3-1-4 Radiographie                                 | <b>11</b> |
| 3-1-5 Imagerie par Résonance Magnétique (IRM)      | <b>11</b> |

|   |           |
|---|-----------|
| 3-2-5 Acquisition des images IRM                                | 12        |
| <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 Approche frontière (contour)                                | 14        |
| 2-1-1-Les méthodes dérivatives                                  | 15        |
| 2-1-2-Méthodes déformables                                      | 15        |
| 2-2-Approche Région   | 16        |
| 2-2-1- Croissance de région (région growing)                    | 16        |
| 2-2-2-Segmentation par fusion de régions (Merge)                | 17        |
| 2-2-3-Segmentation par division de régions (Split)              | 17        |
| 2-2-4-Segmentation par division-fusion (Split and Merge)        | 17        |
| 2-3-Segmentation par classification                             | 18        |
| 2-3-1-Méthodes non Supervisées                                  | 18        |
| 2-3-2-Méthodes supervisées                                      | 18        |
| 2-3-3-Méthode d'ensemble de classifieurs                        | 19        |
| 3- Quelques difficultés de la segmentation des images médicales | 19        |
| Conclusion  | 21        |
| <b>Chapitre2: Apprentissage Automatique</b>                     |           |
| 1- Introduction   | 22        |
| 2 - L'apprentissage automatique                                 | 22        |
| 2- 1-Définition   | 22        |
| 2-2 -Modélisation   | 22        |
| 3 – Domaines de l'apprentissage automatique                     | 23        |
| 3-1- La fouille de données et ses méthodes                      | 23        |
| a - Méthodes descriptives                                       | 24        |
| b-Méthode prédictives   | 25-       |
| 3-2 -L'intelligence Artificielle                                | 25        |
| 4-Types de l'apprentissage automatique                          | 26        |
| 4-1-Apprentissage supervisé                                     | 26        |
| 4-2.Apprentissage non supervisé                                 | 27        |
| 4-3-Apprentissage semi-supervisé                                | 28        |
| 4-4Apprentissage par renforcement                               | 28        |
| 5-Les principaux algorithmes d'apprentissage automatique        | 29        |
| 5-1-Non paramétrique  | 29        |

|   |    |
|---|----|
| 5-1-1 -Les k plus proches voisins   | 29 |
| 5-1-2 Les arbres de décision  | 30 |
| 5-2 Paramétrie  | 31 |
| 5-2-1-Classification naïve bayésienne   | 32 |
| 6- Classification non-supervisée par k-mens                                       | 32 |
| 7-Conclusion:   | 33 |
| <b>Chapitre 3 : Moyennes mobiles locaux pour la segmentation d'IRM cérébrales</b> |    |
| 1- Introduction   | 34 |
| 2- Problématique  | 34 |
| 3- Algorithme k-means   | 35 |
| 3-1 Définition  | 37 |
| 3-2 Objectif de K-means   | 36 |
| 3-3 Principe de la méthode des K-Means  | 36 |
| 3-4- Principe Algorithmique   | 38 |
| 3-5- Organigramme de l'algorithme de k-means                                      | 39 |
| 3-6 Cas d'utilisation k-means   | 39 |
| 3-7 Les avantages de l'algorithme   | 40 |
| 3-8 Les inconvénients   | 40 |
| 4- Application K-means sur IRM  | 40 |
| 5- Application K-means par partie   | 41 |
| 6- Conclusion   | 44 |
| <b>Chapitre 4 : Implémentation et Test</b>  |    |
| 1-Introduction  | 45 |
| 2-Le langage Python   | 45 |
| 2.1-A propos du langage Python  | 45 |
| 2.3 Bibliothèques en Python   | 46 |
| 2.4 Principales Bibliothèques utilisés  | 46 |
| 3- Environnement de développement   | 46 |
| 3-1-Matériel utilisé  | 46 |
| 3-2Outils de développement  | 47 |
| 3-2.1 –Google Colab   | 47 |
| 3-2-2 Les offres de Google Colab  | 48 |
| 4-Création de projet et l'explication de l'interface                              | 48 |

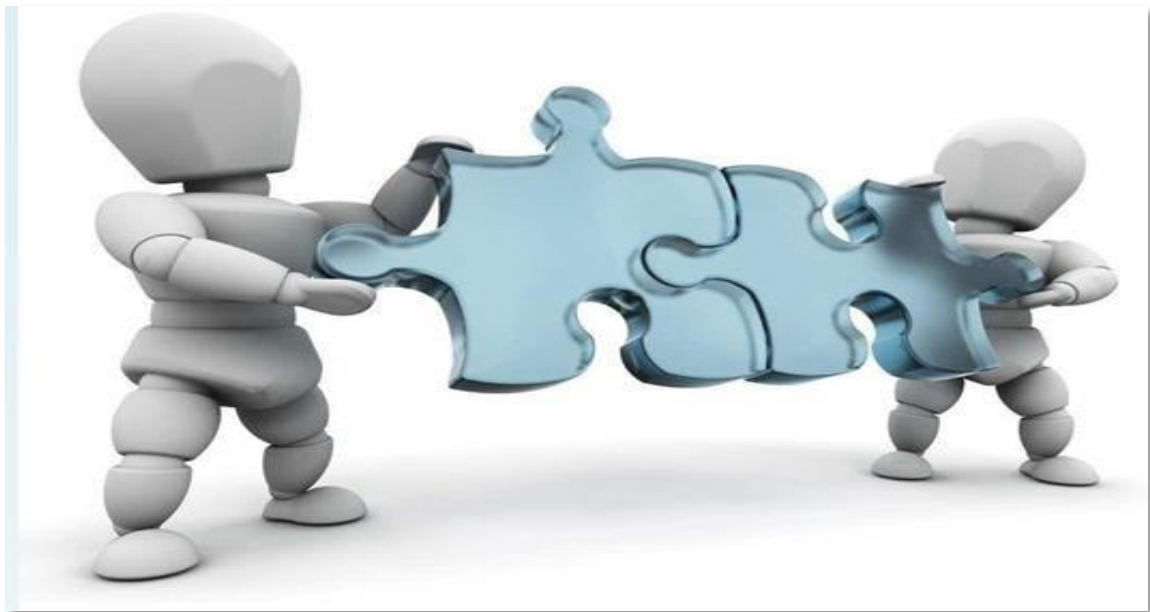
|   |           |
|---|-----------|
| 1-Barre de menus  | 49        |
| 2-Nom de fichier  | 49        |
| 3-Insérer une cellule de code ou texte                                      | 50        |
| 4-Table des matières  | 50        |
| 5-Cellule de texte  | 50        |
| 6-Cellule de code   | 50        |
| 7-Sortie  | 50        |
| 8-Exécuter la cellule   | 50        |
| 9-Fichiers  | 50        |
| 10-Extraits de code   | 50        |
| 11-RAM et disque  | 51        |
| 12-Plus d'options   | 51        |
| 13-Supprimer la cellule   | 51        |
| 14-Miroir la cellule  | 51        |
| 15 Paramètres   | 51        |
| 16-Commentaire  | 51        |
| 17--Lien vers la cellule  | 51        |
| 4-1-Extraits du code  | 51        |
| 4-2-Chargement, filtrage et affichage IRM                                   | 52        |
| 4-3-Chargement données d'apprentissage                                      | 53        |
| 4-4-Initialisation des données d'apprentissage et sélection des sous images | 53        |
| 4-5-K-means local et copie des résultats                                    | 54        |
| 4-6 Affichage du résultat global  | 55        |
| 4-7-Exemple de matrice de confusion   | 55        |
| 4-8-Exemple de Métriques d'évaluation                                       | 55        |
| 4-9-Exemples de résultats de k-means locaux                                 | 56        |
| 4-10-Résultat global  | 57        |
| 5-Conclusion  | 58        |
| <b>Conclusion Générale</b>  | <b>59</b> |
| <b>Bibliographie</b>  | <b>60</b> |

## Table De Figures

| <b>N° Figure</b>          | <b>Designation de figures</b>  | <b>N° Page</b> |
|---------------------------|--|----------------|
| <b>Figures chapitre 1</b> |  |                |
| <b>Figure 1</b>           | Exemple d'image X, Y pixels.   | <b>5</b>       |
| <b>Figure 2</b>           | Image matricielle  | <b>5</b>       |
| <b>Figure 3</b>           | Image vectorielle  | <b>6</b>       |
| <b>Figure 4</b>           | Exemple d'image 12x12 pixels   | <b>7</b>       |
| <b>Figure 5</b>           | Exemple d'image de rayons X  | <b>10</b>      |
| <b>Figure 6</b>           | Exemple d'image Scintigraphies (Médecine Nucléaire)  | <b>10</b>      |
| <b>Figure 7</b>           | Exemple de l'Echographie   | <b>11</b>      |
| <b>Figure 8</b>           | Exemple-de-traces-cephalometriques-realises-sur-des-radiographies-2D-A-Radiographie  | <b>11</b>      |
| <b>Figure 9</b>           | Exemple d'image de résonance magnétique  | <b>12</b>      |
| <b>Figure 10</b>          | Principe de la formation d'IRM   | <b>13</b>      |
| <b>Figure 11</b>          | Les 3 vues IRM selon les 3 principaux axes   | <b>13</b>      |
| <b>Figure12</b>           | Contour et ses dérivées  | <b>15</b>      |
| <b>Figure13</b>           | Segmentation de l'encéphale par croissance de régions [Har05]. (a) segmentation de la MG, (b) segmentation de la MB et (c) régions de chevauchement entre les deux régions segmentées.                               | <b>17</b>      |
| <b>Figure14</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>20</b>      |
| <b>Figures chapitre 2</b> |  |                |
| <b>Figure 1</b>           | Schéma de modélisation d'une machine d'apprentissage   | <b>23</b>      |
| <b>Figure 2</b>           | Schéma des différentes techniques issues de l'IA et FD Pour la construction de modèles de données  | <b>23</b>      |
| <b>Figure 3</b>           | Processus d'extraction de connaissances à partir des données   | <b>24</b>      |
| <b>Figure 4</b>           | Les différents types d'apprentissage automatique   | <b>26</b>      |
| <b>Figure 5</b>           | Schéma d'un modèle supervisé   | <b>27</b>      |
| <b>Figure 6</b>           | Schéma d'un modèle non supervisé   | <b>28</b>      |
| <b>Figure 7</b>           | Schéma d'un modèle semi-supervisé  | <b>28</b>      |

|                           |   |           |
|---------------------------|---|-----------|
| <b>Figure 8</b>           | Schéma d'un modèle par renforcement                     | <b>29</b> |
| <b>Figure 9</b>           | Fonctionnement de kplus proche voisins                  | <b>30</b> |
| <b>Figures chapitre 3</b> |   |           |
| <b>Figure 1</b>           | k-means-clustering-algorithm                            | <b>35</b> |
| <b>Figure 2</b>           | Exemple d'initialisation du K-Means avec trois clusters | <b>36</b> |
| <b>Figure 3</b>           | Mise à jour des centres après une itération de K-means  | <b>37</b> |
| <b>Figure 4</b>           | Exemples de clusters finaux obtenus avec K-means        | <b>37</b> |
| <b>Figure 5</b>           | Organigramme de l'Algorithme K-Means                    | <b>39</b> |
| <b>Figures chapitre 4</b> |   |           |
| <b>Figure 1</b>           | google colab  | <b>47</b> |
| <b>Figure 2</b>           | Création un projet dans Google Colab                    | <b>49</b> |
| <b>Figure 3</b>           | Partie gauche de l'interface                            | <b>49</b> |
| <b>Figure 4</b>           | Partie droite de l'interface                            | <b>51</b> |
| <b>Figure 5</b>           | Image brute à segmenter                                 | <b>53</b> |
| <b>Figure 6</b>           | Exemple des k-means locaux                              | <b>56</b> |
| <b>Figure 7</b>           | Image global segmentée par k-means                      | <b>57</b> |

# Introduction Général



## Introduction générale

Le domaine de l'imagerie englobe un large éventail d'applications. L'imagerie a connu un essor considérable en raison des potentialités qu'elle offre. Dans le domaine médical, elle est devenue un élément clé du quotidien des médecins (diagnostic, assistance aux interventions chirurgicales, etc.). Chaque jour, diverses applications produisent des quantités énormes d'images médicales. Un seul examen par scanner CT (Tomographie Computée) peut générer des centaines d'images. Ces images sont appelées images médicales volumétriques ou simplement images volumétriques. Traiter et exploiter ces images est une tâche complexe. La segmentation automatique des images médicales issues de divers examens courants est cruciale car elle aide le praticien à prendre des décisions diagnostiques. L'efficacité de cet outil repose sur deux facteurs principaux : le temps de calcul et la qualité de segmentation (E.-H. Guerrou, Mahiou, & Ait-Aoudia, 2013a; E. L. H. Guerrou, Ait-Aoudia, Michelucci, & Mahiou, 2017).

La segmentation est une des étapes critiques de l'analyse d'images. Elle permet d'isoler les objets pertinents dans une image pour les analyser. Ces objets peuvent ensuite être manipulés plus aisément dans diverses applications telles que la reconnaissance d'objets en imagerie médicale et traitement de documents, l'indexation (recherche d'un objet dans une base d'images en biométrie), la compression, le recalage d'images, etc. Pour cela, les chercheurs ont développé de nombreuses méthodes de segmentation en fonction du domaine d'application (Cerqueira et al., 2002; Dale, Fischl, & Sereno, 1999; Girshick, Donahue, Darrell, & Malik, 2014; Liberman, Shankweiler, Fischer, & Carter, 1974).

En imagerie par résonance magnétique (IRM), le problème de non-uniformité de l'intensité (INU : Intensity Non Uniformity), qui consiste en une variation d'intensité due au dispositif physique d'acquisition, est un défi supplémentaire en plus du 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 cette méthode consiste à classer les pixels de l'image en travaillant dans une région locale où le problème d'INU est minimisé et peut être négligé. Ensuite, les résultats obtenus dans les différentes régions locales sont fusionnés.

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

Ce mémoire est structuré comme suit :

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

**Le chapitre 2 :** est consacré aux méthodes d'apprentissage automatique, en nous concentrant sur les méthodes supervisées, et en particulier l'algorithme des moyennes mobiles (k-means).

**Le chapitre 3 :** détaille notre méthode, en expliquant son principe et ses différentes étapes.

Nous y montrons comment les régions locales sont générées et comment les classes fiables de pixels sont calculées.

**Le chapitre 4 :** traite de l'implémentation et des tests de la méthode proposée, où nous présentons l'environnement et le langage de programmation utilisé, à savoir Python. Nous y montrons quelques interfaces de l'application et discutons de certains résultats expérimentaux.

**Une conclusion générale :** résume notre travail et en souligne les perspectives.

# CHAPITRE 01

Segmentation D'images

Médicales

## 1-Introduction :

L'imagerie médicale est un domaine essentiel de la médecine qui utilise différentes technologies pour visualiser l'intérieur du corps humain. Ces techniques permettent aux médecins de diagnostiquer les maladies, d'évaluer les lésions et de suivre l'évolution des traitements. Parmi les principales modalités d'imagerie médicale, on trouve la radiographie, l'échographie, l'IRM (imagerie par résonance magnétique), la TDM (tomodensitométrie) et la scintigraphie. Chaque technique a ses propres applications et avantages, et souvent plusieurs modalités sont combinées pour obtenir un diagnostic plus complet. L'imagerie médicale a révolutionné la pratique médicale en permettant aux professionnels de la santé d'avoir un regard non invasif et détaillé à l'intérieur du corps humain.

La segmentation médicale est le processus de délimitation et d'identification des structures anatomiques ou des régions d'intérêt dans les images médicales. Cela peut inclure la segmentation des organes, des tissus, des lésions ou des anomalies présentes dans les images. La segmentation est une étape essentielle dans l'analyse d'images médicales car elle permet d'extraire des informations quantitatives sur les structures anatomiques, d'aider à la planification chirurgicale, de quantifier la progression des maladies, etc.

Dans ce chapitre, La première partie, nous présentons une notion générale sur les images et l'imagerie médicale. 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 les différentes méthodes de la segmentation d'images médicales IRM, et à la fin de ce chapitre on discutera sur les différents problèmes de segmentation.

## Partie 1 : Imagerie Médicale

### 2- Notion générale sur les images :

#### 2-1 Définition d'image :

L'image est une reproduction significative d'ensemble d'information structurée caractérisant un univers composé d'entités : objets figurant dans une scène réelle, la sculpture, le dessin, la photographie, le film, ...etc.

# CHAPITRE1: Segmentation d'Images médicale

C'est aussi un ensemble structuré d'informations qui a une signification pour l'œil humain, après l'affichage sur l'écran [1].

On peut être décrit l'image sous deux formes qui sont :

- Image numérique
- Image analogique

## a. Image analogique

Elle est sous la forme d'une fonction  $I(X, Y)$  de brillance analogique continue, définie dans un domaine borné, tel que  $X$  et  $Y$  sont les coordonnées spatiales d'un point de l'image et une fonction d'intensité lumineuse et de couleur. Sous cette forme, l'image est inexploitable par l'ordinateur qui ne connaît que le langage binaire, ce qui nécessite sa numérisation.

La numérisation d'une image analogique est la conversion de celle-ci de son état analogique (distribution continue d'intensités lumineuses), en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques  $I(x, y)$  ou :

$x, y$  : Coordonnées cartésiennes d'un point de l'image.

$I(x, y)$  : niveau de gris en ce point.

L'opération de numérisation comprend trois phases :

\_ **Échantillonnage** : l'échantillonnage commence par découper l'image en surfaces carrées élémentaires d'une matrice carrée ou rectangulaire ; chacun des carrés élémentaires d'image est appelé *pixel* (Picture Cell) et repère par ses coordonnées  $x$  et  $y$ .

\_ **Quantification** : les lignes sont étudiées les unes après les autres et sur chacune, la valeur de chaque pixel est mesurée ; parfois ce pixel a une structure hétérogène ; la valeur retenue est alors une moyenne, des détails seront donc perdus.

\_ **Codage** : représentation informatique (binaire) des valeurs représentant les pixels.[2]

## b. Image numérique :

L'image numérique est un ensemble de données divisée en éléments de taille fixe appelée pixel ayant chacun une valeur de couleurs correspondante à l'image réelle et numériser sur trois octets de trois couleurs RVB (rouge, vert, bleu) ou sur un octet dans le cas d'une image de niveau de gris. Cette numérisation facilite la transformation, le stockage et le traitement de l'image.

On peut définir une image comme une fonction  $I(x, y)$  où  $x$  et  $y$  sont des coordonnées spatiales,  $I(x, y)$  est le pixel de coordonnées  $x$  et  $y$  comme il est illustré dans la figure 1. [3].



**Figure 1** : Exemple d'image X , Y pixels.

## 2-2 Types d'images

### 1. Image matricielle (bitmap)

Une image matricielle est une image en mode point. Le système de codage le plus universel consiste en effet à décomposer la représentation graphique, l'image, en un certain nombre de points élémentaires dit pixel, ces pixels sont caractérisés par leurs coordonnées spatiales et leur couleur [4].



**Figure 2** : Image matricielle

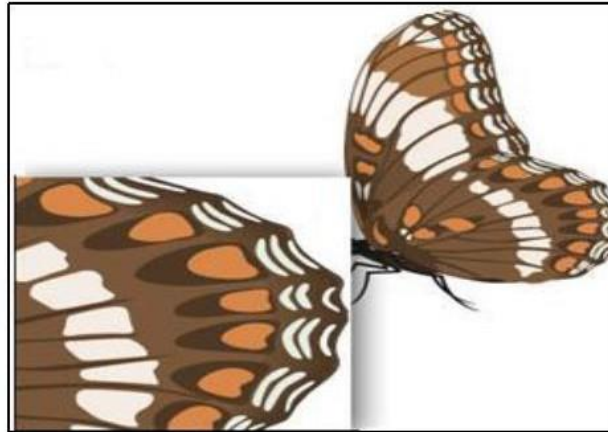
### 2. Image vectorielle

Dans une image vectorielle les données sont représentées par des formes géométriques simples qui sont décrites d'un point de vue mathématique. Il s'agit de représenter les données de l'image par des formules géométriques qui vont pouvoir être décrites d'une façon mathématique.

# CHAPITRE1: Segmentation d'Images médicale

Autrement dit, on stocke la succession d'opérations conduisant au tracé dans le cas d'une image vectorielle, alors qu'on mémorise une mosaïque de points élémentaires dans le cas d'image matricielle.

Ces images présentent deux avantages : elles occupent peu de place en mémoire et peuvent être redimensionnées sans perte d'information [4]



**Figure 3** : Image vectorielle

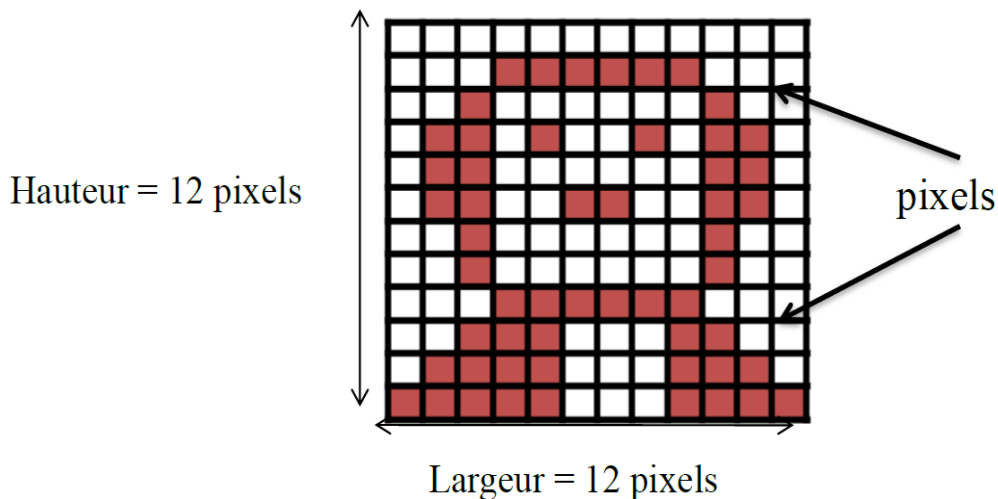
## Caractéristique d'une image numérique

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants :

### 1. Pixel :

Une image est constituée d'un ensemble de points appelés pixels, contraction de l'expression anglaise « Picture elements » éléments d'image. Le pixel représente ainsi le plus petit élément constitutif d'une image numérique, c'est une entité calculable qui peut recevoir une structure et une quantification. L'ensemble de ces pixels est contenu dans un tableau à deux dimensions constituant l'image.

La quantité d'information que véhicule chaque pixel donne des nuances entre les images monochromes et les images couleurs. Dans les images monochromes, chaque pixel est codé sur un octet. Et dans une image couleur (R, V, B), un pixel peut être représenté sur trois octets, un octet pour chacune des couleurs : rouge(R), vert(V), bleu(B).



**Figure 4 : Exemple d'image 12x12 pixels.**

## **2. Bruit :**

Le bruit est très souvent présent sur les images de documents, car il peut apparaître à différents endroits de la chaîne de numérisation : à l'impression, pendant la vie du document et à la numérisation.

Des traitements classiques tels que les filtres médians sont utilisés pour filtrer des pixels isolés sur l'image. Les opérateurs de morphologie mathématique sont également très souvent utilisés pour corriger de légers défauts sur les images comme "recoller" des morceaux de caractères.

## **3. Dimension :**

La dimension est la taille de l'image, se présente sous forme de matrice dont les éléments sont des valeurs numériques représentative des intensités lumineuses (pixel), le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre totale de pixels dans une image.

## **4. Le contraste :**

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images.

Si  $L_1$  et  $L_2$  sont les degrés de luminosité respectivement de deux zones voisines  $A_1$  et  $A_2$  d'une image, le contraste  $C$  est défini par le rapport suivant :

$$C = \frac{L1 - L2}{L1 + L2}$$

## 5. Résolution :

Le mot résolution est la taille d'une image (nombre de pixels), utilisé pour désigner le nombre totale de pixels affichables horizontalement ou verticalement, où ce nombre est plus grand, la résolution est meilleure.

## 6. Extension :

L'extension d'un fichier, c'est le point, et les trois ou quatre lettres qui suivent le nom d'un fichier. Grâce à ces quelques lettres, vous pouvez connaître le type du fichier (application, documents texte, page web, image, etc). Il suffit d'un clic droit sur le fichier et de lire les propriétés pour trouver l'extension de ce fichier.

- Exemple d'extension :

- Word.exe = application

- Document.txt = document texte

- Index.htm = page web

- Image.jpg (.gif) = image

## 7. La luminance :

La luminance est tout simplement la sensation visuelle qu'on a de la lumière. De toutes les grandeurs qui caractérisent la lumière, la luminance est celle qui nous concerne le plus, lorsque la vision est gênée dans une forte pénombre, ou lorsqu'un phare nous éblouit, c'est la luminance qui est en cause.

Il y a plusieurs facteurs qui influencent le niveau de luminance. En premier bien sûr, l'intensité lumineuse de la source de lumière. Ensuite lorsque la luminance est renvoyée par une surface éclairée. Et enfin, la coloration de la source lumineuse influence aussi le niveau de luminance [1].

## 3. L'imagerie Médicale :

L'imagerie médicale est certainement l'un des domaines de la médecine qui a le plus progressé ces dernières années. Elle regroupe l'ensemble des techniques et les moyens d'acquisition et de restauration l'image 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.

### 3-1-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 :

#### 3-1-1 Scanographie :

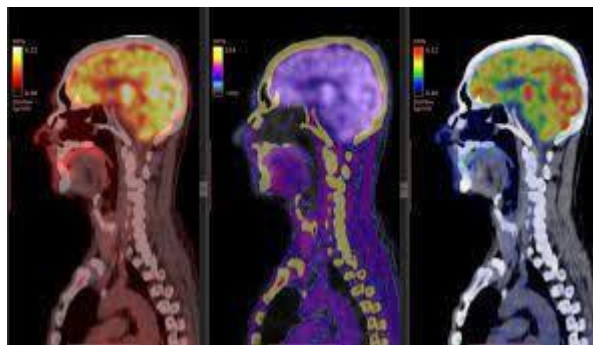
La scanographie et la tomodensitométrie émettent des faisceaux de rayons X autour du patient ce qui va permettre, grâce à l'ordinateur, de reconstituer très précisément l'organe examiné. Cette technique a permis de grandes avancées, notamment dans l'exploration du cerveau, la recherche de tumeurs, la visualisation de métastases et les études des traumatismes.



**Figure 5 :** Exemple d'image de rayons X

### 3-1-2 Scintigraphies (Médecine Nucléaire)

Les images de la médecine nucléaire représentent la distribution spatiale au sein d'un organe d'un isotope radioactif fixé à un traceur injecté dans l'organisme. Cet ensemble marqueur-traceur a un devenir physiopathologique sans toxicité et sans perturbation des équilibres métaboliques. Cet ensemble permet diverses investigations non invasives des fonctions biologiques biochimiques ou biophysiques des organes, selon les traceurs utilisés.



**Figure 6 :** Exemple d'image Scintigraphies (Médecine Nucléaire)

### 3-1-3 Echographie

L'échographie est une technique d'exploration de l'intérieur du corps basée sur les ultrasons. Elle consiste à envoyer sur une région du corps, au moyen d'une sonde, des faisceaux d'ultrasons par brèves impulsions successives. L'impulsion ultrasonore réfléchiée par les tissus rencontrés constitue l'écho. Les échos recueillis par la sonde fourniront, après amplification, un signal vidéo. Selon les types de sondes et les techniques, on obtient des images de formes sectorielles ou rectangulaires. L'échographie, reconnue inoffensive, trouve son principal champ d'application dans l'observation des tissus mous et la surveillance de la grossesse.



Figure 7 : Exemple de l'Echographie

## 3-1-4 Radiographie

Découverte il y a plus d'un siècle, la radiographie utilise les rayons X, capables de traverser la matière. Passant à travers une certaine partie du corps, ils impressionnent un film radiographique, plus ou moins noirci en fonction de l'organe traversé. La "radio" ressemble ainsi à une ombre chinoise, où les os apparaissent en blanc et les structures moins denses (comme les poumons) en noir. Ces films doivent être numérisés pour être traités informatiquement.

L'opération de numérisation peut être réalisée en utilisant soit un système à base de négatoscope avec une caméra vidéo, soit avec un scanner de films radiologiques.

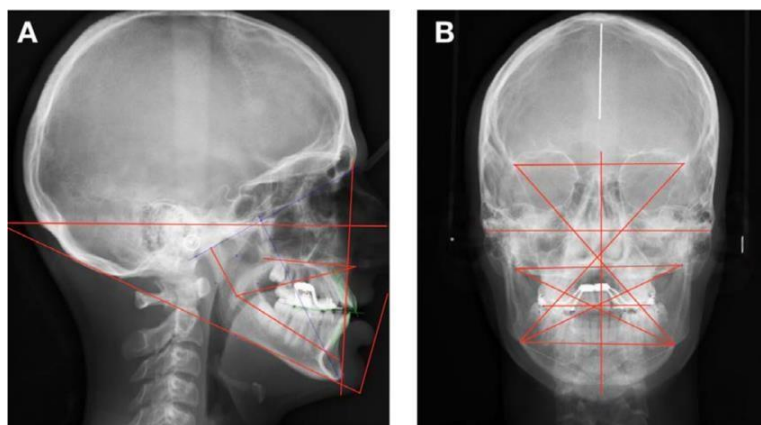


Figure 8 : Exemple-de-traces-cephalometriques-realises-sur-des-radiographies-2D-A-Radiographie

## 3-1-5 Imagerie par Résonance Magnétique (IRM)

# CHAPITRE1: Segmentation d'Images médicale

L'imagerie par résonance magnétique permet d'analyser des organes de manière très précise. Elle permet de visualiser des détails invisibles sur la radiographie standard, l'échographie ou le scanner. Par exemple elle recherche au niveau du cerveau, des lésions infectieuses ou inflammatoires, des anomalies des vaisseaux, ainsi que des tumeurs.

La technique repose sur l'utilisation d'un aimant qui va générer un champ magnétique ayant une action sur les ions hydrogènes contenus dans l'eau du corps. Les données recueillies sont ensuite traitées par informatique, ce qui permet d'obtenir des images de qualité en 2D ou après reconstruction informatique, en 3D. Cette technique fournit en même temps des informations non seulement sur la morphologie et la physique de l'organe, mais aussi sur les réactions biochimiques qui s'y passent.

Ainsi les médecins peuvent connaître avec précision l'état d'un organe et suivre, sans risque pour le patient, l'évolution d'un traitement pharmaceutique ou les suites d'une opération. Les résultats de l'IRM sont remarquables, le contraste des tissus dépasse celui des meilleurs scanners, et tous les plans de coupe sont possibles (le scanner n'autorise que des coupes horizontales dites axiales ou transversales) [5].

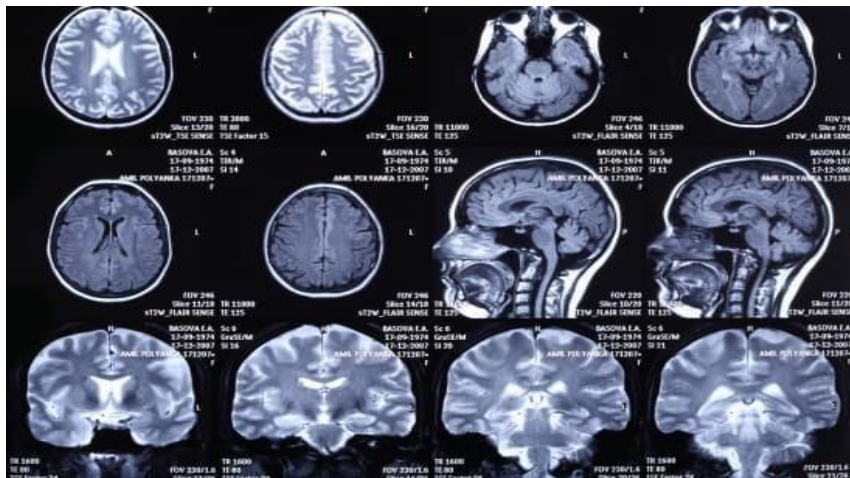


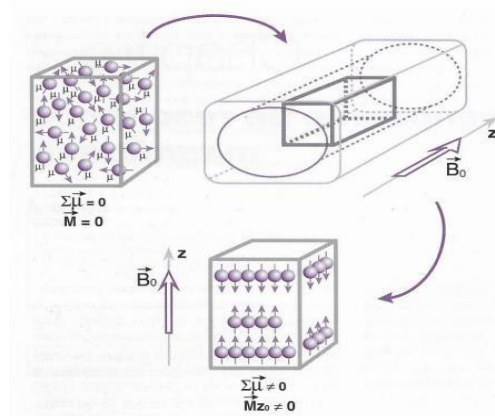
Figure 9 : Exemple d'image de résonance magnétique

## 3-2-5Acquisition des images IRM:

La formation d'images (données) IRM résulte de l'enregistrement de la résonance des atomes d'hydrogènes, contenus dans l'eau abondante dans le corps humain, et qui entrent dans la composition des tissus des organes cibles [6]. De ce fait, plus la concentration de l'eau est élevée dans les tissus, plus la qualité de l'IRM obtenue est la meilleure. Le procédé technologique permettant d'obtenir une IRM consiste à faire appliquer sur l'organisme un champ magnétique puissant, ce qui induit l'unification de la polarisation des molécules d'eau, qui se dirigent toutes

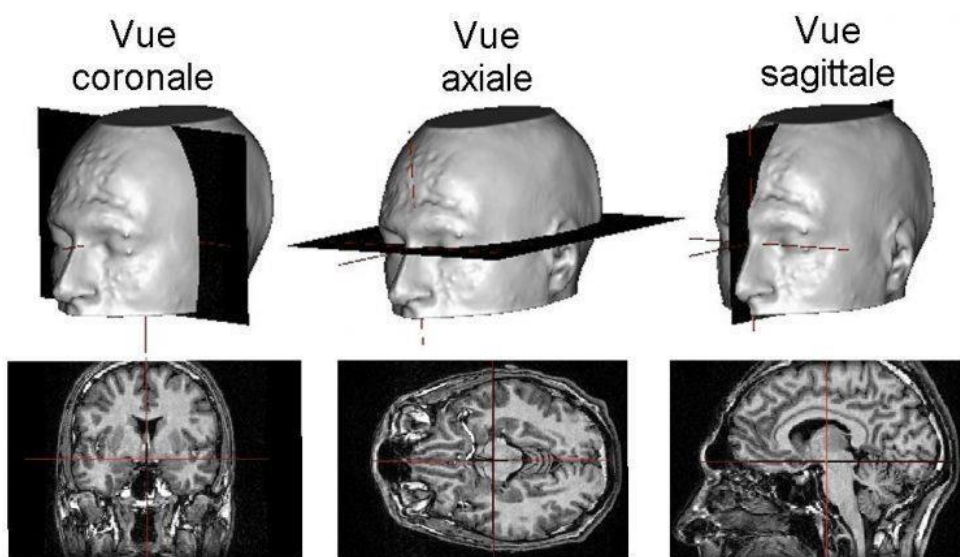
# CHAPITRE1: Segmentation d'Images médicale

dans la direction du champ (Figure 2.1). Le retour à l'état initial des molécules d'eau engendre un signal qui sera reçu et traité par un dispositif informatique pour former le volume de données IRM.



**Figure 10** :Principe de la formation d'IRM

La technique IRM permet donc la visualisation des tissus mous de l'organisme contenant de l'eau. Ces tissus émettent des signaux d'intensités différentes ce qui permet d'avoir des écarts de contraste après leur représentation. Le principe de l'IRM diffère de celui du scanner, car cette dernière technique est basée sur l'absorption des tissus des rayons 'X' qui traversent l'organisme. Etant donné que la masse de données obtenue par une IRM est un volume tridimensionnel de forme rectangulaire 3D, il est possible de la visualiser selon les 3 principaux axes qui forment le référentiel 3D. On obtient trois vues nommées respectivement coronale, axiale et sagittale (voir Figure 2).



**Figure 11** :Les 3 vues IRM selon les 3 principaux axes

## Partie 2 : Segmentation d'image

### Définition de la segmentation

La segmentation est un traitement de bas-niveau qui consiste à effectuer une partition de l'image en régions homogènes, connexes par rapport à un ou plusieurs critères. Les régions obtenues se distinguent les unes des autres par des différences significatives selon ces mêmes critères. Après ces étapes, nous pouvons introduire un traitement sectoriel de différentes manières.

La segmentation consiste à extraire des points, des lignes ou des régions. Le choix d'une technique de la segmentation est liée à plusieurs facteurs comme : la nature de l'image, les conditions d'acquisition (bruit), les primitives à extraire (contour, texture,...).

La segmentation fait référence aux notions de similarité comme les perçoit, le système visuel humain et ceci donne naissance à deux approches couramment qualifiées d'approche « Région » et d'approche « frontière ».

Il existe plusieurs catégories de méthodes pour réaliser la segmentation d'image :

- Segmentation par région.
- Segmentation par contour.
- Segmentation par seuillage.
- Segmentation par coopération région-contour

### 2- Les différentes méthodes de segmentation :

#### 2-1 Approche frontière (contour) :

Dans une image numérique, les contours se situent entre les pixels appartenant à des régions ayant des intensités moyennes différentes ; il s'agit de contours de type « saut d'amplitude ». Un contour peut également correspondre à une variation locale d'intensités présentant un maximum ou un minimum ; il s'agit alors de contour « en toit ». La notion de « frontière » est associée à une variation d'intensité ou à une discontinuité entre les propriétés de deux ensembles connexes de points [07].

Donc l'approche locale de détection de contour consiste à balayer l'image avec une fenêtre définissant la zone d'intérêt. A chaque position, un opérateur est appliqué sur les pixels de la fenêtre afin d'estimer s'il y a une transition significative au niveau de l'attribut choisi. A partir des pixels susceptibles d'appartenir à un contour, il faut ensuite extraire des contours fermés [08].

#### 2-1-1-Les méthodes dérivatives :

# CHAPITRE1: Segmentation d'Images médicale

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 [9].

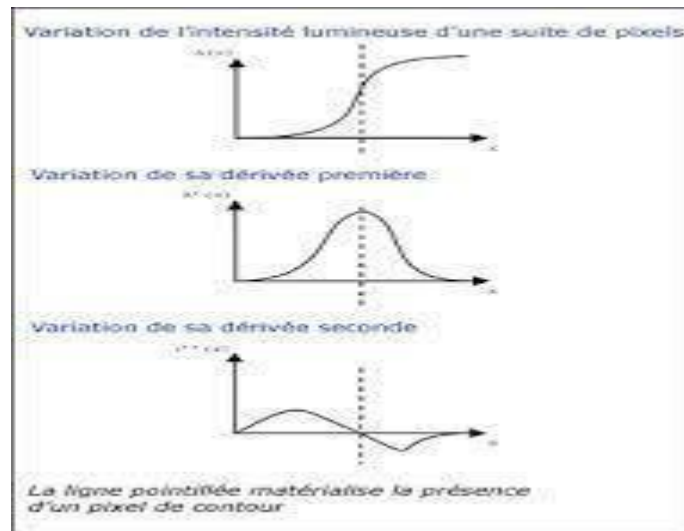


Figure 12: Contour et ses dérivées.

## 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 [10], 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 [11] :

- 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.

# CHAPITRE1: Segmentation d'Images médicale

- Il est souvent difficile d'identifier et de classer les contours parasites.

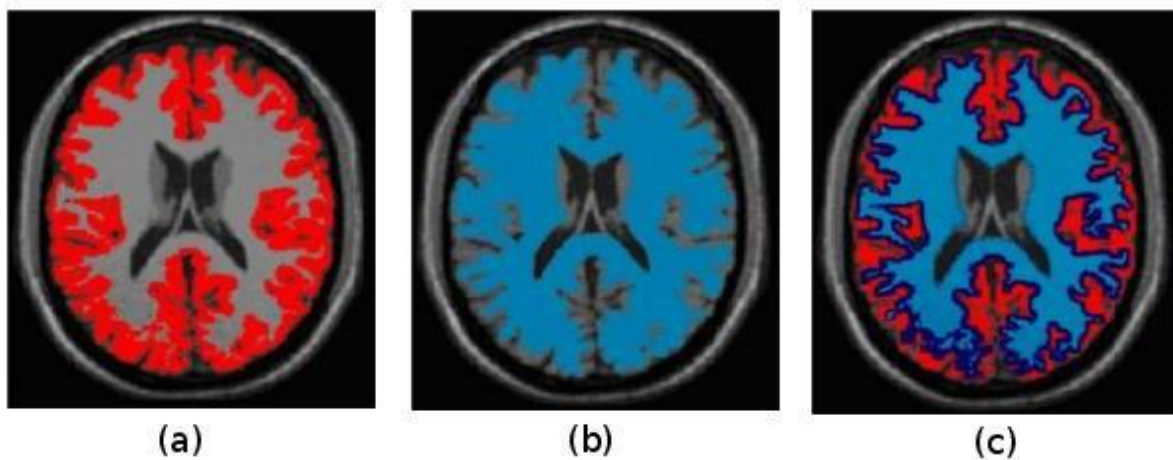
## 2-2-Approche Région :

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 Donnée. 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 [11]. 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 [12].



# CHAPITRE1: Segmentation d'Images médicale

**Figure 13:** Segmentation de l'encéphale par croissance de régions [Har05].  
(a) segmentation de la MG, (b) segmentation de la MB et (c) régions de chevauchement entre les deux régions segmentées.

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

Les techniques de réunion (regionmerging) 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 [12].

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 [10].

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.

## 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 d'adjacence de régions .On s'arrête quand le critère de fusion n'est plus vérifié [9].

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.

# CHAPITRE1: Segmentation d'Images médicale

□ 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.

### **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 [13]. 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 [14].

### 3- Quelques difficultés de la segmentation des images médicales :

#### ➤ **Bruit :**

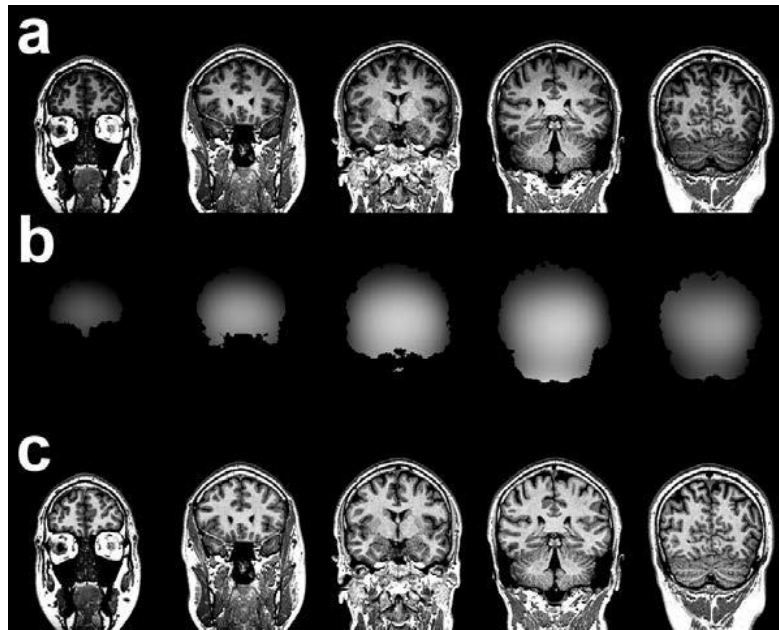
Le bruit d'image caractérise la présence d'informations parasites qui s'ajoutent de façon aléatoire aux détails dans les images acquises. Il est plus particulièrement visible dans les zones peu éclairées, où le rapport signal/bruit est faible. Il a pour conséquence la perte de netteté dans les détails [15].

#### ➤ **Inhomogénéités du signal dans le volume :**

Il est techniquement très difficile de maintenir un champ magnétique spatialement constant. Le champ magnétique radiofréquence  $B_1$  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.

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 [16].



**Figure 14: 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.**

## ➤ 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[16].

## ➤ Artéfact de mouvement

L'artéfact de mouvement est un des artéfacts les plus fréquemment rencontré. Il se constitue lorsqu'il y a translation dans l'espace du segment étudié au cours de l'acquisition. Il y a deux types de mouvements rencontrés [15]:

# CHAPITRE1: Segmentation d'Images médicale

- 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.

## Conclusion

En combinant l'imagerie médicale avec des techniques de segmentations avancées, les cliniciens et les chercheurs peuvent obtenir des informations précieuses pour améliorer la compréhension des pathologies, personnaliser les traitements et améliorer les résultats pour les patients. La segmentation médicale peut être réalisée manuellement par des experts en imagerie médicale, mais elle est également de plus en plus automatisée grâce à des algorithmes de traitement d'image et d'apprentissage automatique, ce qui permet d'accélérer et de standardiser le processus.

# CHAPITRE 02

## 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.

## 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[17].

### 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.



Figure 1 : Schéma de la 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 [18].

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

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

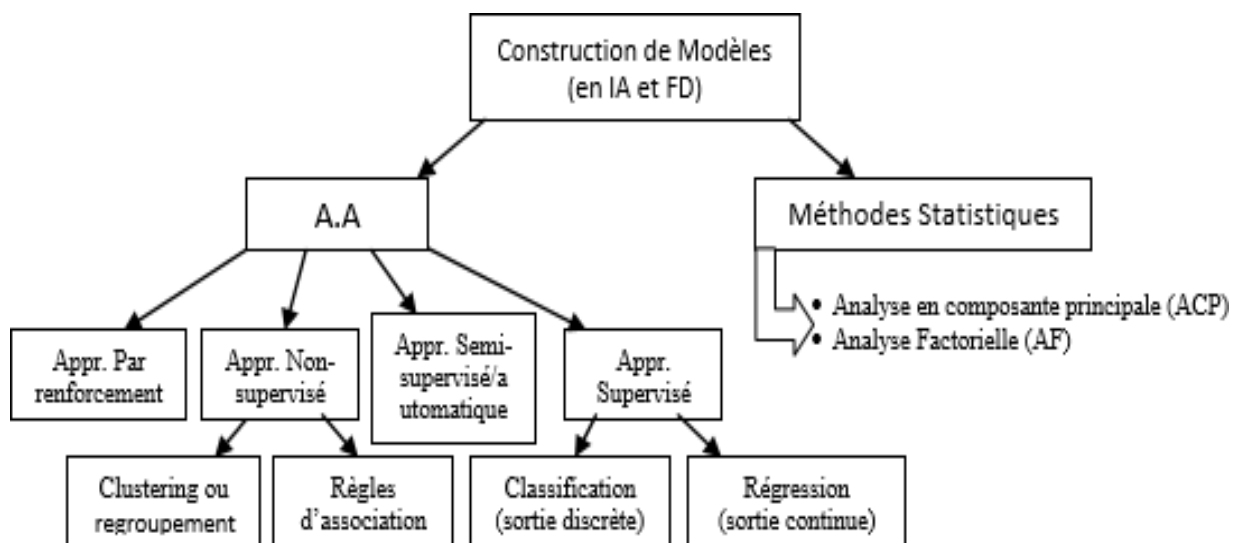


Figure 2 : Schéma des différentes techniques issues de l'IA et FD pour la construction de modèles de données

#### 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 ») [19].

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[20].

Le processus du data Mining "fouille de données" représente une étape d'un processus plus grand : l'ECD. D'après [21], 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.

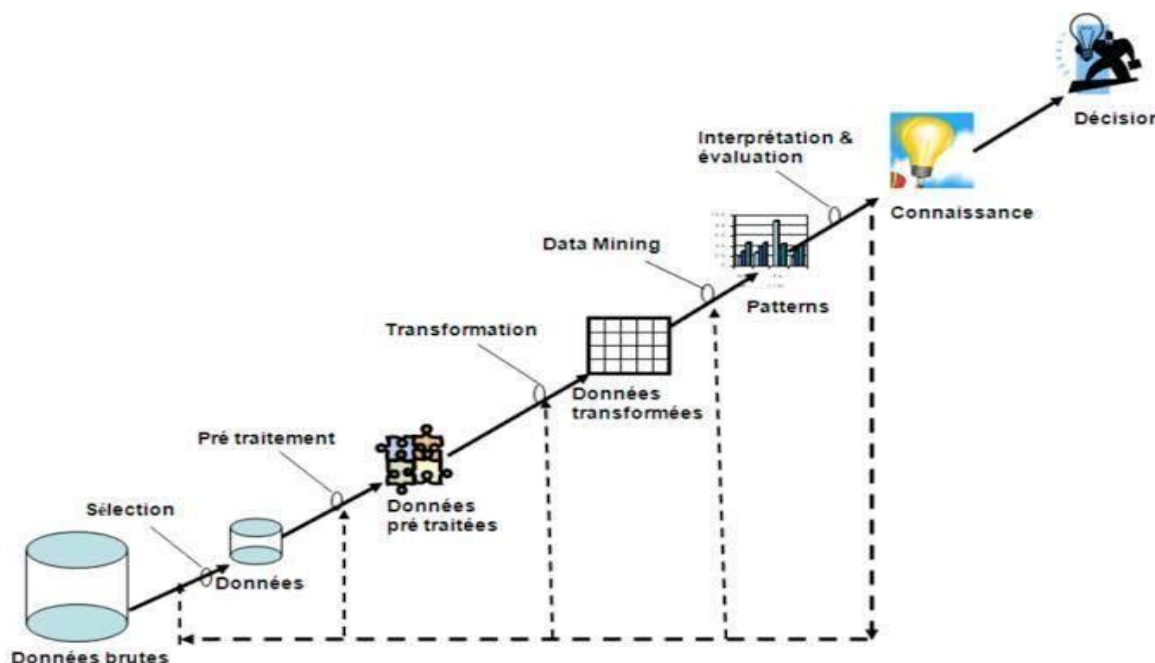


Figure 3 : Processus d'extraction de connaissances à partir des données

Pour arriver à exploiter ces quantités importantes de données, le data Mining 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 [22]. 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. [23].

### ➤ **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 [23].

### **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 dépendant ou client à budget[22].

- **Scoring**

Prédiction de valeur inconnu à partir de valeur connu .exemple: accordé oui ou non un prêt à un client [22].

### **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 [24].

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 à leur spécificité et son utilisé en fonction du problème que nous rencontrons..

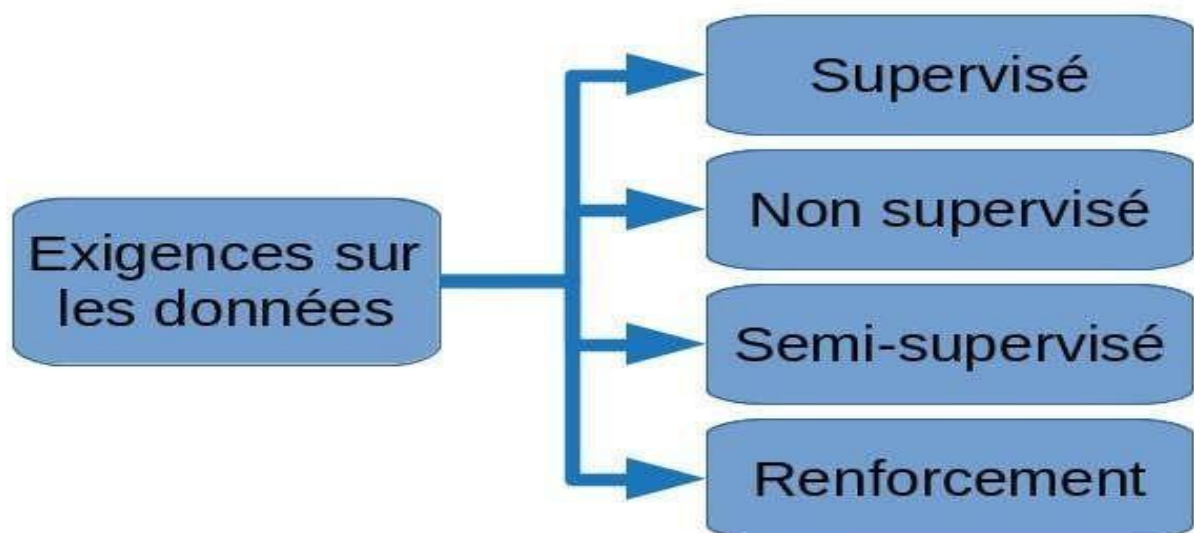


Figure 4: Les différents types d'apprentissage automatique

#### 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 [25].

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 5 : Schéma d'un modèle supervisé

### 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 [25].

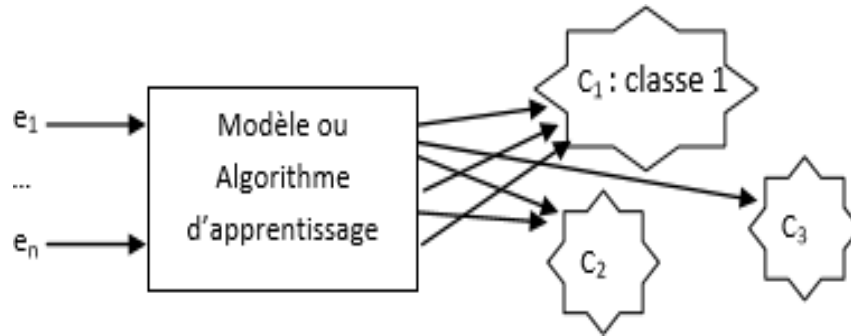


Figure 6: Schéma d'un modèle non supervisé

### 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 [25].



Figure 7: 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 [25].

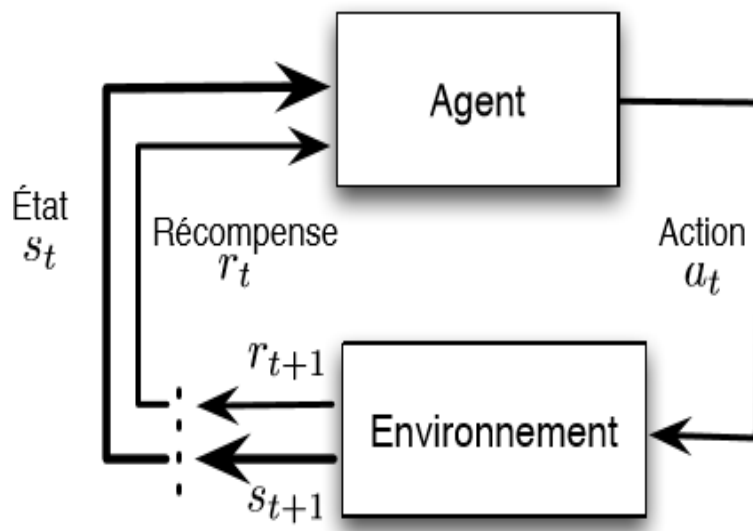


Figure 8: Schéma d'un modèle par renforcement

### 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 hypothèse sur le modèle que suivent les données. C'est le cas des plus proches voisins (k-PPV), arbre de décision, [26].

##### 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 qualifié 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 que l'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.

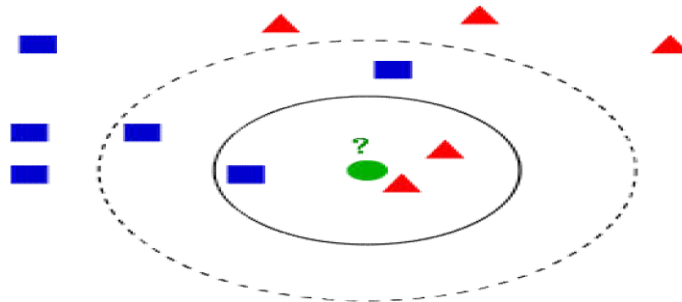


Figure 9 : Fonctionnement de  $k$  plus proche voisins

**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 [27].

### 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.

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 [28].

### 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 [29].

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

La méthode de classification naïve bayésienne est un algorithme d'apprentissage supervisé (*supervised machine learning*) qui permet de classifier 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 donnée, 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 soit connu et fournit, d'où le caractère supervisé de l'outil [30].

### 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.

## 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.

### **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

## 1- Introduction :

Le partitionnement (Clustering) est une tâche importante en analyse de données, elle divise un ensemble de données en plusieurs sous-ensembles, ces sous-ensembles appelés groupes ou clusters. [31]

Dans ce cadre plusieurs méthodes ont été développées, nous détaillerons une de ces méthodes, la plus populaire en l'occurrence la méthode des k moyennes (K-means) mise au point par McQueen en 1967, elle doit sa popularité à sa simplicité et sa capacité de traiter de larges ensembles de données. [32]

Nous présenterons tout d'abord une brève description du fonctionnement de l'algorithme k-means. Finalement, nous avons appliqué cette dernière en une série d'exemples testés à la fin de ce chapitre.

## 2- Problématique :

La segmentation d'IRM cérébrale pose plusieurs défis, et une problématique clé pourrait être : "Comment améliorer la précision et la robustesse de la segmentation d'IRM cérébrale pour permettre un diagnostic plus précis des pathologies neurologiques ?"

En abordant ces défis, on peut développer des algorithmes de segmentation plus avancés, des techniques de prétraitement d'image améliorées, ainsi que des approches d'apprentissage automatique robustes qui prennent en compte la variabilité anatomique et pathologique pour obtenir des résultats plus précis et fiables.

Donc on a proposé 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).

## 3- Algorithme k-means :

### 3-1 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), 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 [33].

Contrairement à d'autres méthodes dites hiérarchiques, qui créent une structure en « arbre de clusters » pour décrire les groupements, 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.

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. [33]

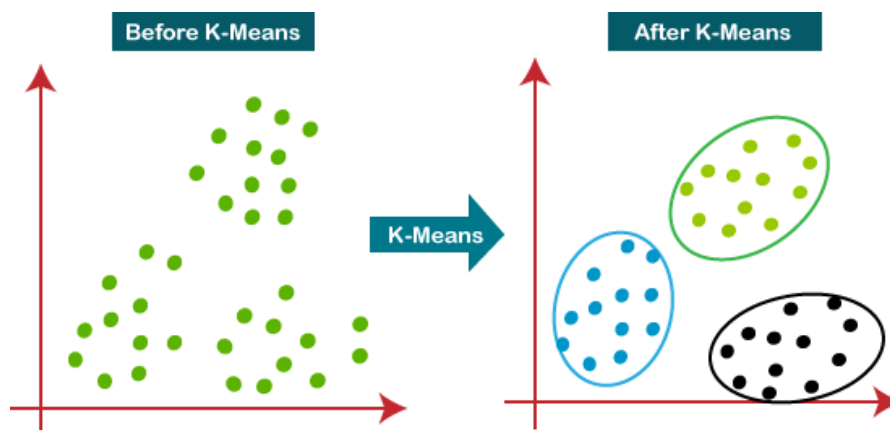


Figure 1: k-means-clustering-algorithm

## 3-2 Objective de K-means:

La méthode des "K-means" reste actuellement la méthode la plus utilisée dans la classification des images surtout pour les grands fichiers de données qui contiennent plus de 40 000 individus. En effet, cette méthode a été utilisée pour classer 40 000 personnes. Ceux-ci ont répondu à une enquête sur les ventes par correspondance d'une entreprise afin d'obtenir des profils types de clientèle. Elle procède par le calcul des distances entre les objets et les k centroïdes et on affecte les objets aux centroïdes dont ils sont les plus proches. Elle effectue en suite un tri et enfin elle agrège les individus les plus proches. [34]

## 3-3 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. [35]

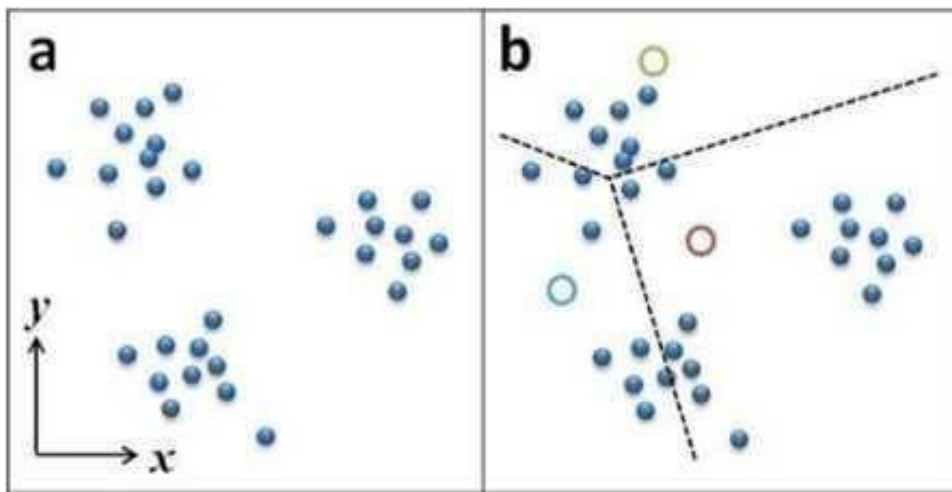
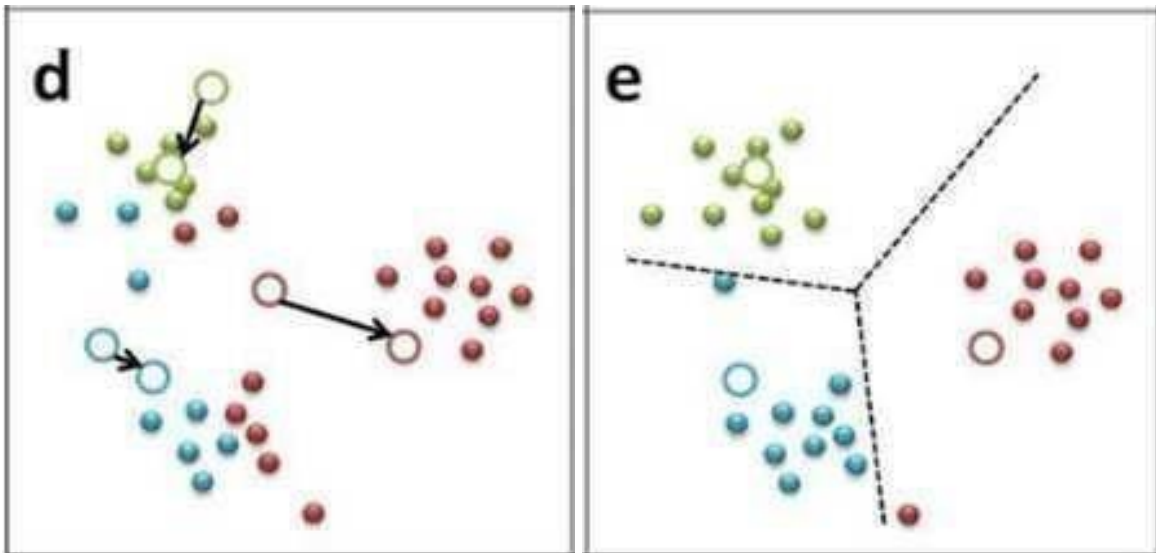


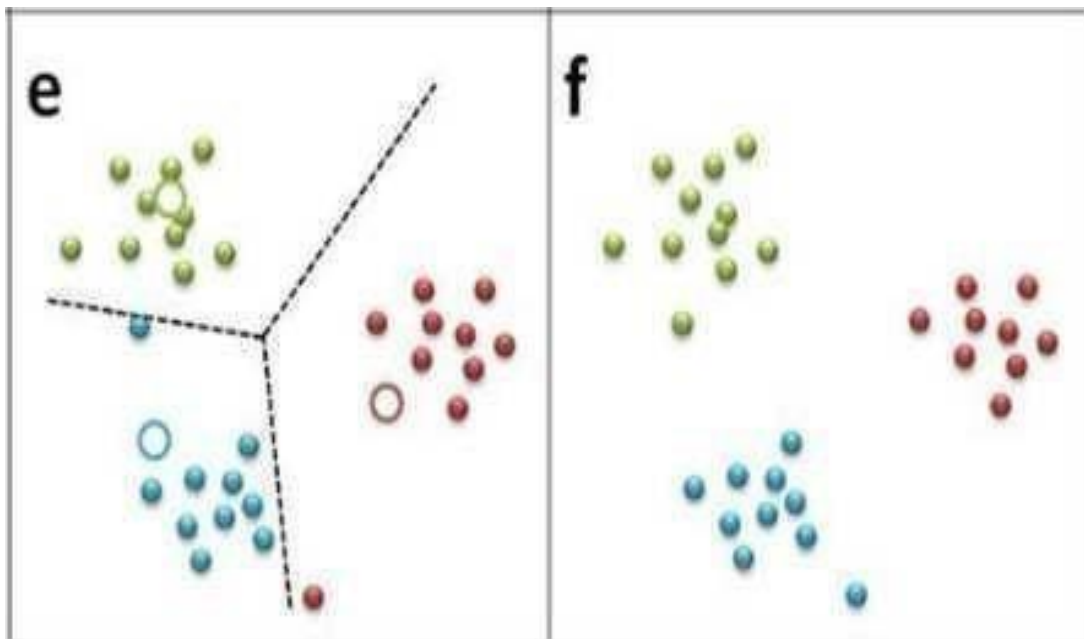
Figure 2 : Exemple d'initialisation du K-Means avec trois clusters

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

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.

### 3-4- Principe Algorithmique : [36]

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;

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

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

Fin.

## 3-5- Organigramme de l'algorithme de k-means:

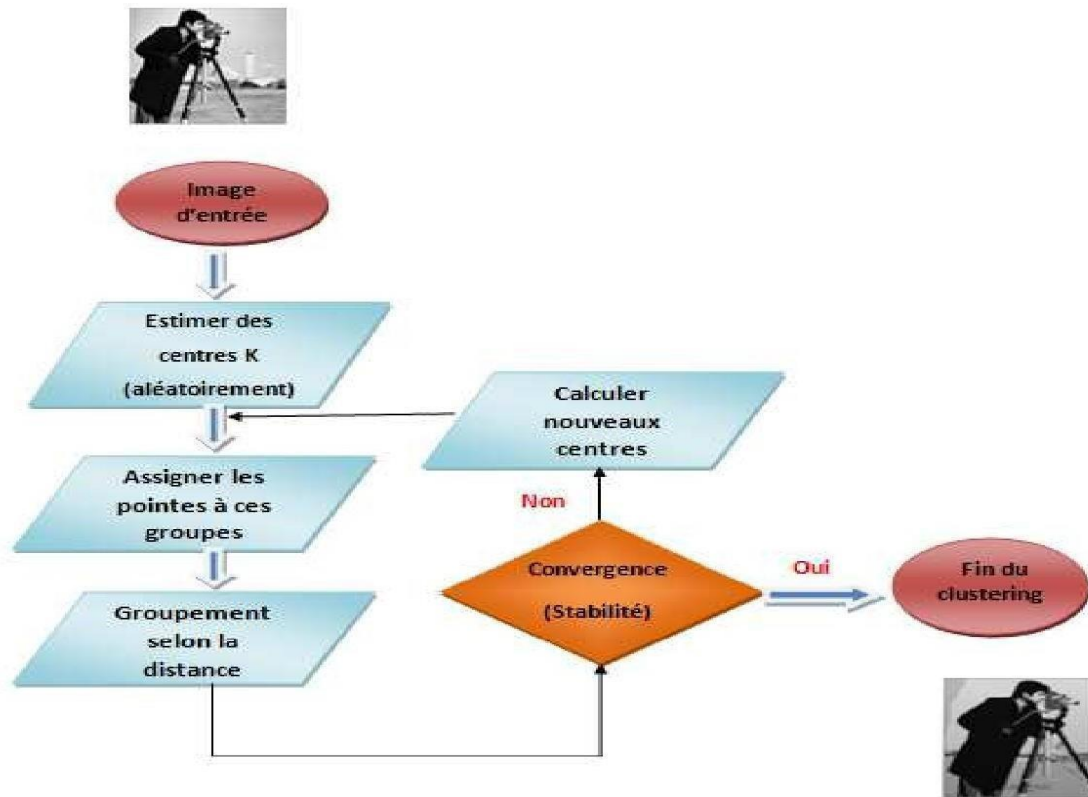


Figure 5 : Organigramme de l'Algorithme K-Means

## 3-6 Cas d'utilisation k-means :

K-Means en particulier et les algorithmes de clustering de façon générale ont tous 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.) [37]

### 3-7 Les avantages de l'algorithme :

- Son efficacité et sa simplicité sont en fonction de la forme des clusters. Les K-Means fonctionnent bien dans les clusters hyper-sphériques.
- Applicable à des données de grandes tailles, et aussi à tout type de données (mêmes textuelles), en choisissant une bonne notion de distance.
- Comparée à l'utilisation d'autres méthodes de classification, une technique de classification k-means est rapide et efficace en termes de coût de calcul.
- Les résultats sont très faciles à interpréter. K-Means génère des descriptions de cluster sous une forme minimisée pour maximiser la compréhension des données[38] .[39]

### 3-8 Les inconvénients :

- 1) Le nombre de classe doit être fixé au départ,
- 2) Le résultat dépend de tirage initial des centres des classes,
- 3) Les clusters sont construits par rapports à des objets inexistants (les milieux).[39]

## 4- Application K-means sur IRM :

L'application de l'algorithme k-means sur les images par résonance magnétique (IRM) est une technique couramment utilisée en neuroimagerie pour segmenter et classifier différentes régions anatomiques du cerveau en fonction de leurs propriétés de signal. Les différentes catégories que vous mentionnez - le fond, la matière grise, la matière blanche et le liquide céphalorachidien (LCR) - représentent des composants anatomiques distincts du cerveau qui peuvent être différenciés à partir des données IRM.

Voici comment l'algorithme k-means pourrait être appliqué à ces données :

1. **Prétraitement des données** : Les images IRM doivent être prétraitées pour améliorer la qualité et faciliter la segmentation. Cela peut inclure des étapes telles que la correction du mouvement, la normalisation d'intensité, et la correction du champ magnétique.
2. **Extraction de caractéristiques** : Les caractéristiques pertinentes sont extraites des images IRM pour représenter les différentes régions anatomiques. Cela peut inclure des caractéristiques de texture, d'intensité, de forme, etc.

3. **Choix du nombre de clusters** : Pour appliquer k-means, il faut déterminer le nombre optimal de clusters, c'est-à-dire le nombre de régions anatomiques distinctes que l'on souhaite segmenter. Des techniques comme la validation croisée ou les critères d'information peuvent être utilisées pour cela.
4. **Segmentation avec k-means** : Une fois que le nombre optimal de clusters est choisi, l'algorithme k-means est appliqué aux caractéristiques extraites des images IRM pour segmenter les différentes régions anatomiques en clusters.
5. **Post-traitement et visualisation** : Après la segmentation, des techniques de post-traitement peuvent être appliquées pour améliorer la précision et la cohérence des résultats. Enfin, les résultats peuvent être visualisés pour évaluer la qualité de la segmentation et interpréter les différentes régions anatomiques identifiées.

Il convient de noter que l'application de k-means aux données IRM peut être complexe et nécessite une expertise en traitement d'images et en neuroimagerie pour obtenir des résultats précis et significatifs. De plus, d'autres algorithmes de segmentation, tels que les méthodes basées sur les contours actifs ou les réseaux de neurones convolutifs, peuvent également être utilisés en combinaison avec k-means pour améliorer la précision de la segmentation.

### 5- Application K-means par partie

Nous proposons ci-dessous un algorithme permettant l'application de Kmeans par partie dans l'IRM à segmenter. L'idée est de varier à chaque itération la taille de la sous image, dans laquelle le Kmeans est appliquée. Ceci permet de diversifier les classes d'un pixel donné. Cependant une classe remplace une autre, seulement si l'écart-type des intensités autour du pixel en question est inférieur à l'écart type de la sous-image sélectionnée, et ce pour en assurer que les pixels auront à la fin les étiquettes les plus pertinentes et donc les plus correctes.

Nous procédons d'abord à un lissage de l'IRM en utilisant un lissage adaptatif. Ce dernier est exécuté pour réduire le bruit mais préserver en même temps les détails pertinentes dans l'images, notamment au niveau des pixels situés sur les contours entre les régions.

### Algorithme Lissage\_Adaptatif

**Entree** : irm brute de taille WxH

**Sortie** irmL lissée de taille WxH

#### Début

Pour y allant de 1 à H-2 faire

  Pour x allant de 1 à W -2 faire

    S = 0

    NbrVoisins = 0

    Pour i allant de -1 à 1 faire

      Pour j allant de -1 à 1 faire

        Si  $\text{abs}(\text{irm}[y+i][x+j]-\text{irm}[y][x]) < \text{seuil}$  alors

          S=S+irm[y+i][x+j]

          NbrVoisins = NbrVoisins +1

        Finsi

      irmL[y][x]= S/NbrVoisins

    Fin pour

  Fin pour

#### Fin

La valeur du seuil est choisie de sorte que les pixels voisines appartenant à des régions différentes ne soient pas lissés.

L'algorithme du calcul du k-means local itère sur l'étiquetage des pixels de l'irm lissée en utilisant des sous images de cette dernière.

### Algorithme Kmean\_par\_partie

**Entree** irm brute

**Sortie** irm segmentée

#### Début

Pour itération allant de 1 à 10 faire

    xd = random(1,taille\_rectangle)

    xf = random(W-taille\_rectangle, W-1)

    yd = random(1,taille\_rectangle)

    yf = random(H-taille\_rectangle, H-1)

    X\_Apprentissage = sous\_irm(xd,yd,xf,yf)

    Kmeans(X\_Apprentissage, nbr\_classes = 4)

    Y\_predict = kmeans.etiquette

        Pour y allant de yd à yf faire

            Pour x allant de xd à xf faire

                Si  $\text{ecarttype}(\text{voisinage}(x,y)) < \text{ecarttype}(xd,yd,xf,yf)$  alors

                    Irmsegmentée[y][x] = Y\_predict[y][x]

                Fin si

            Fin pour

        Fin pou

Fin pour

**Fin.**

Pour chaque itération, l'algorithme choisit une sous image dans le rectangle (xd,yd)-(xf,yf), puis procède à l'exécution de l'algorithme de Kmeans sur ce rectangle. Le résultat obtenu (Y\_Predict) est copié sur le rectangle correspondant dans l'image segmentée. Cependant une étiquette de pixel n'est remplacée par la nouvelle étiquette dans Y\_Predict que si l'écart-type des intensités des pixels dans le voisinage du pixel en question est inférieure à l'écart-type des intensités des pixels de la sous image (xd,yd)-(xf,yf). Ceci permet une sélection automatique des étiquettes les plus pertinentes des pixels dans une région donnée.

### **6- 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 la segmentation partielle appartenant aux différentes régions locales ce qui préserve les pixels de contours de ne pas être étiquetés comme pixels de l'intérieur des régions.

Au chapitre suivant, nous présenterons l'implémentation et l'expérimentation de la méthode proposée.

# CHAPITRE 04

## 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 [40].

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 [40].

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.

## 2.2 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. [41]

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. [41]

## 2.3 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.

## 3. Environnement de développement :

### 3.1-Matériel utilisé :

Notre travail s'est effectué sur Labtop 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. [42].



**Figure 1 : google colab**

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 [43].

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 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 [43].

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) :

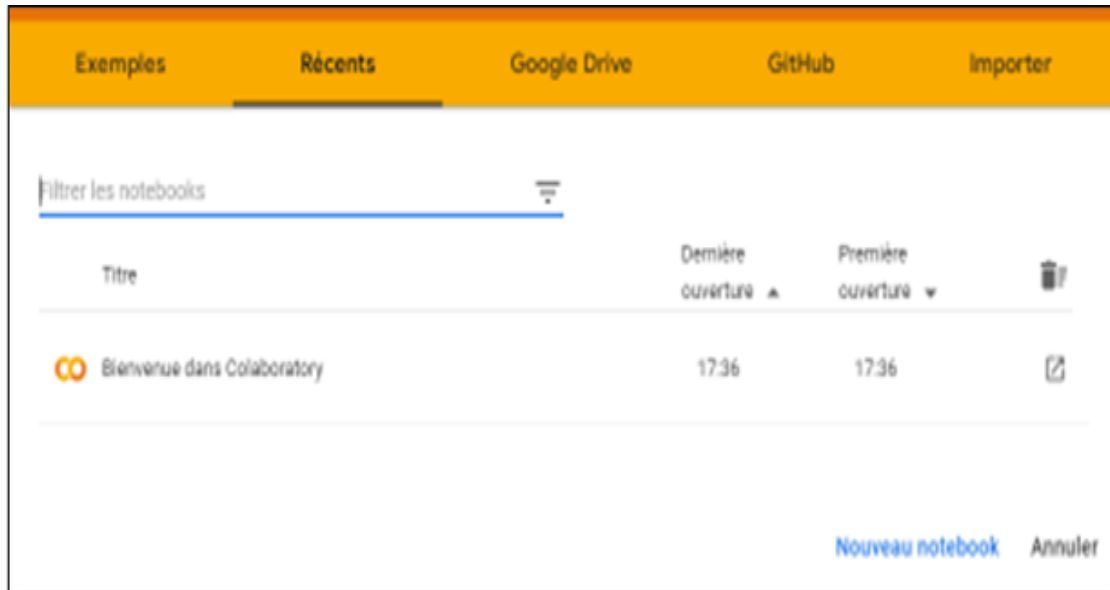


Figure 2 : Création un projet dans Google Colab

**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

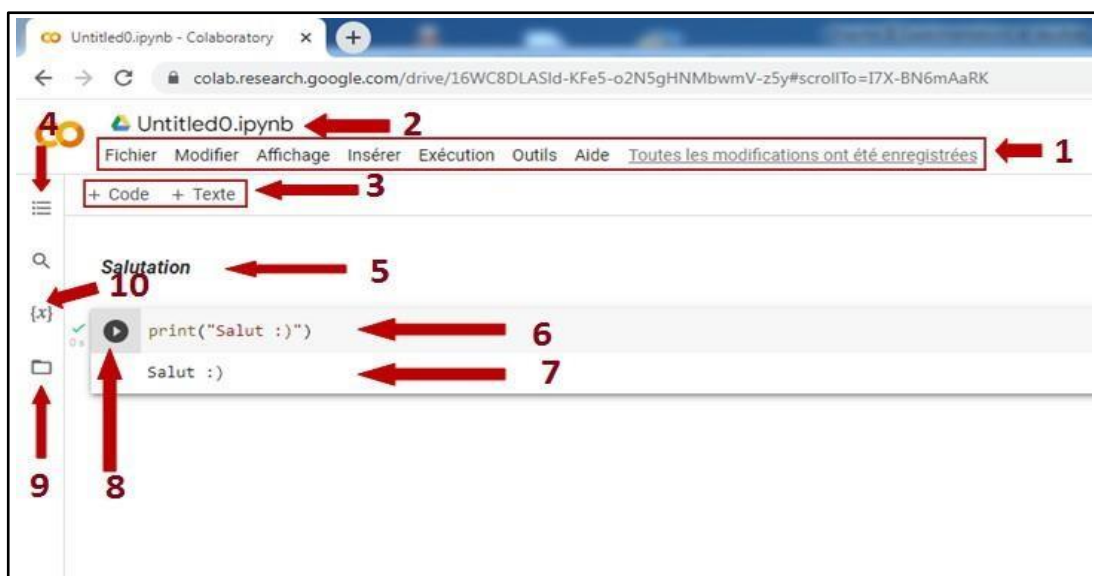


Figure 3 : Partie gauche de l'interface

**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

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.



Figure 4 : 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.

### 4-1-Extraits du code :

Modules utilisés

```
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*
from PIL import Image
from matplotlib import pyplot as plt
import numpy as np
```

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

```
❖ #chargement de l'IRM (coupe N° 83)
❖ #-----
❖ H = 217
❖ L = 181
❖ IRMfile="/content/drive/MyDrive/Slice83.rawb"
❖ fichier=open(IRMfile,"rb")
❖ image = bytearray(fichier.read())
❖ fichier.close()
❖
❖ Amélioration IRM par filtre adaptatif
❖ #-----
❖ def IRM(x,y):
❖     return image[y*L+x]
❖ for y in range(1,H-1):
❖     for x in range(1,L-1):
❖         s=0
❖         Nbvoisins=0
❖         for yp in range(y-1,y+2):
❖             for xp in range(x-1,x+2):
❖                 if (abs(IRM(xp,yp)-IRM(x,y))<20):
❖                     Nbvoisins=Nbvoisins+1
❖                     s=s+IRM(x,y)
❖         image[y*L+x]=s//Nbvoisins
❖
❖ #Affichage de l'IRM
❖ #-----
❖ print("Image avant traitement")
❖ frame=Image.new('RGB',(181,217))
❖ for y in range(H):
❖     for x in range(L):
❖         c= image[y*L+x]
❖         frame.putpixel((x,y),(c,c,c))
❖ plt.figure()
❖ plt.imshow(frame)
❖
```

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

```
#Chargement IRM d'entrainement
#-----
filename="/content/drive/MyDrive/rt83.rawb"
fichier=open(filename,"rb")
rt = bytearray(fichier.read())
fichier.close()
```

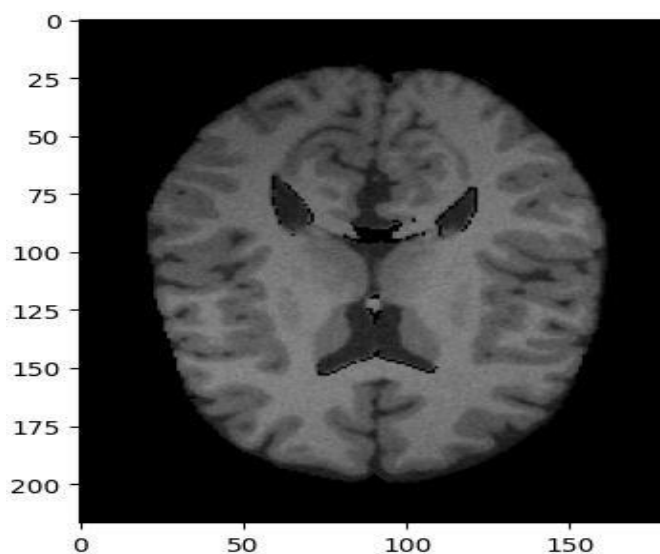


Figure 5 : Image brute à segmenter

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

```
# sauvegarde des images initiales
# -----
ImInit=np.array(image)
RtInit=np.array(rt)
# matrice global des étiquettes
ImTot=np.zeros(L*H)

#Itérations sur kmeans
for iteration in range(10):
    xd=randint(1,70)
    xf=randint(L-70,L-1)
    yd=randint(0,70)
    yf=randint(H-70,H-1)
    if(iteration == 9):
        xd,xf,yd,yf=0,L-1,0,H-1
```

```
print(xd, xf, yd, yf)

# Préparation de la sous image IRM
X = []
for y in range(yd, yf+1):
    for x in range(xd, xf+1):
        X.append(ImInit[L*(y-1)+x])

Y = []
for y in range(yd, yf+1):
    for x in range(xd, xf+1):
        Y.append(RtInit[L*(y-1)+x])

X=np.array(X)
Y=np.array(Y)

L2 = xf-xd+1
H2 = yf-yd+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).

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

```
# k-means local
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)
# Transferte l'étiquette sur la matrice totale
for y in range(H2):
    for x in range(L2):
        ImTot[L*(y+yd)+x+xd]= Y_predict[y*L2+x]

# sous image avec k-means local
print("Sous Image étiquetée par kmeans")
imkmeans=Image.new('RGB', (L2, H2))
for y in range(H2):
    for x in range(L2):
        c= 30*Y_predict[y*L2+x]
        imkmeans.putpixel((x, y), (c, c, c))
plt.figure()
plt.imshow(imkmeans)

# fin boucle sur kmeans
```

```
# -----
```



## 4-6-Affichage du résultat global :

```
❖ #Affichage résultat global
❖ print("Image gobale segmentée par kmeans")
❖ imkmeansglob=Image.new('RGB', (L,H))
❖ for y in range(H):
❖     for x in range(L):
❖         c= int(30*ImTot[y*L+x])
❖         imkmeansglob.putpixel((x,y), (c,c,c))
❖ 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.

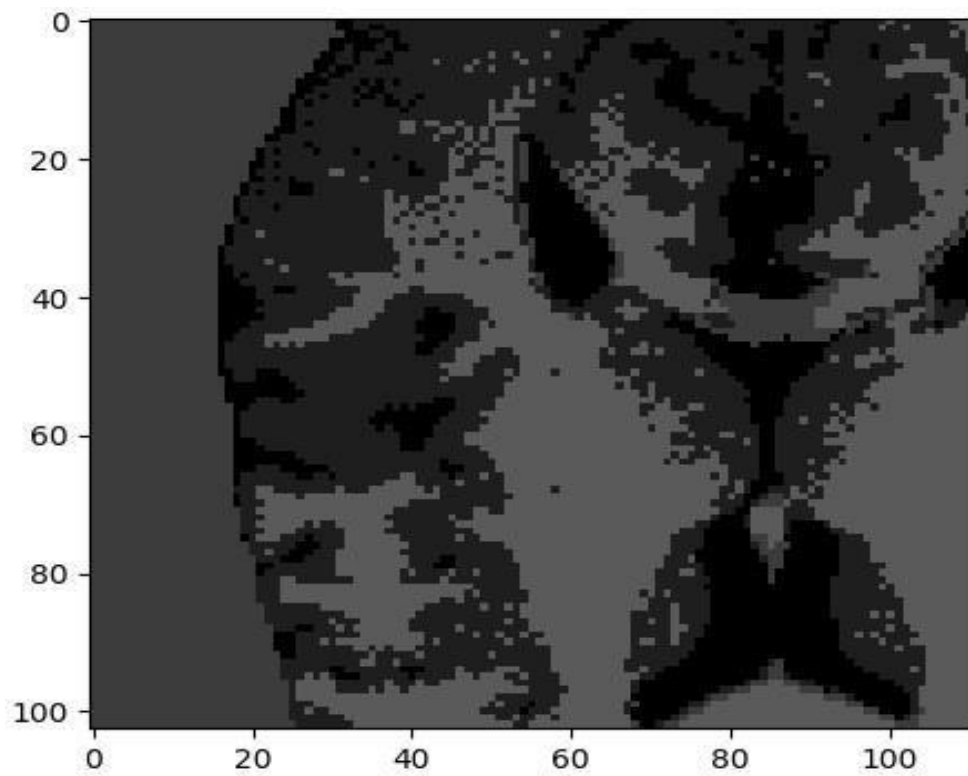
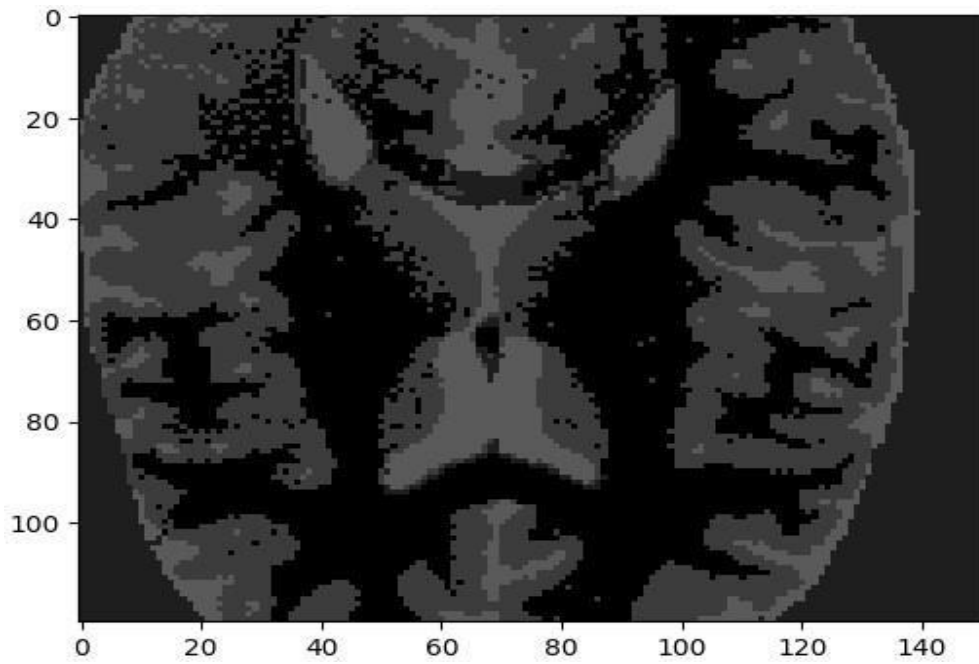
## 4-7-Exemple de matrice de confusion :

```
[[ 0  0  0 1308]
 [ 0 1646 172  26]
 [ 627 691 5409  0]
 [5202  8 1609  0]]
```

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

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.00      | 0.00   | 0.00     | 1308    |
| 1            | 0.70      | 0.89   | 0.79     | 1844    |
| 2            | 0.75      | 0.80   | 0.78     | 6727    |
| 3            | 0.00      | 0.00   | 0.00     | 6819    |
| accuracy     |           |        | 0.42     | 16698   |
| macro avg    | 0.36      | 0.42   | 0.39     | 16698   |
| weighted avg | 0.38      | 0.42   | 0.40     | 16698   |

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



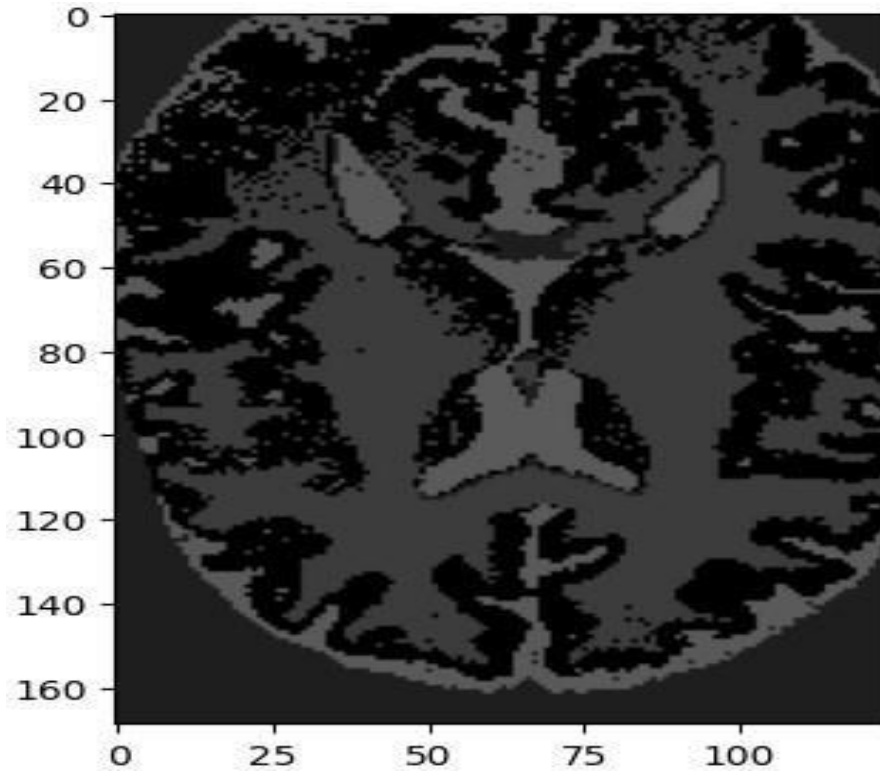


Figure 6 :03 Exemple des k-means locaux

**4-10-Résultat global :**

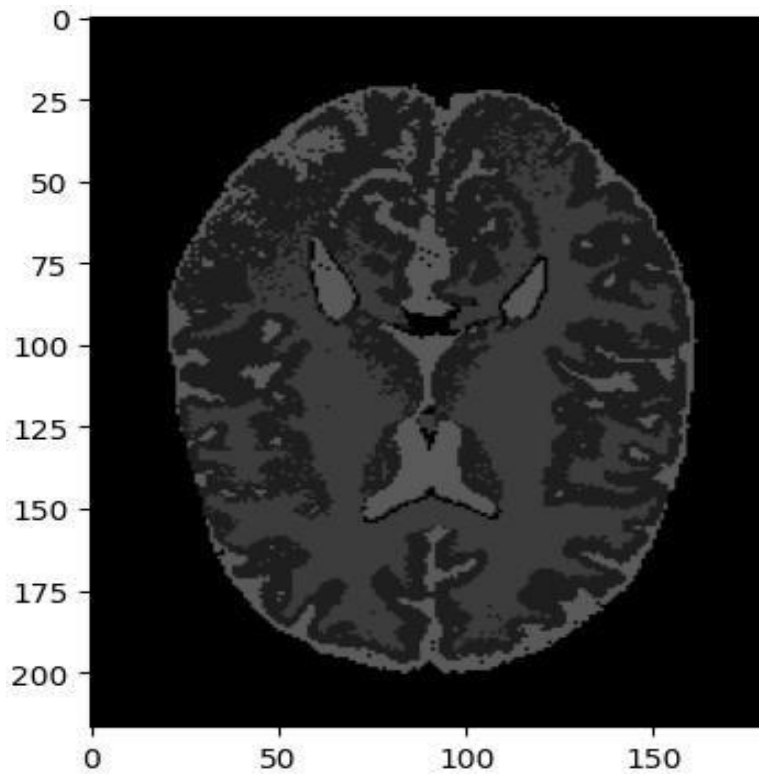
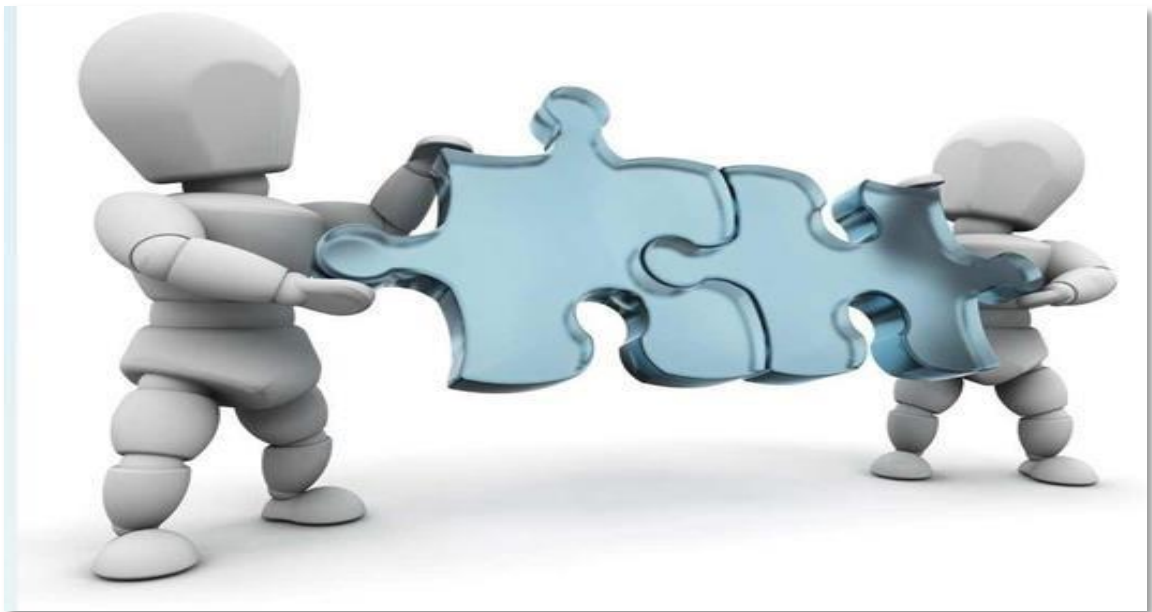


Figure 7 : Image global segmentée par k-means

### **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 particulier les IRM, reste problématique en raison du bruit et des diverses déformations, notamment la non-uniformité des intensités, qui biaisent les résultats de segmentation. Pour améliorer la segmentation des IRM cérébrales, ce mémoire de master propose une nouvelle méthode basée sur un clustering local des données IRM, suivi de la fusion des résultats partiels pour obtenir un clustering global supérieur à celui réalisé sur l'image entière.

Nous avons choisi l'algorithme de clustering k-means, largement utilisé pour la segmentation des IRM. La sélection des régions locales se fait de manière aléatoire, en veillant à ce que les régions soient de petite taille, ce qui permet de surmonter le problème de la non-uniformité des intensités. En effet, dans une petite région locale, ce problème peut être négligé car il n'affecte pas significativement les moyennes des classes.

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

Pour les perspectives de ce travail, il s'agit de quantifier la performance de la méthode proposée et de la comparer à d'autres méthodes. Il serait également intéressant d'utiliser d'autres algorithmes de clustering, tels que le DBSCAN.

# *Bibliographie*

## **Bibliographie**

- [1] -Mémoire pour l'obtention de diplôme de master « Détection d'objet basé sur l'estimation de la saillance » Présenté par : Melle Katia DJAILEB et Melle Wahiba OUALLI « Université Mouloud Mammeri De Tizi-Ouzou 2017/2018 ».
- [2]-Mémoire pour l'obtention de diplôme de l'ingénieur d'état « Segmentation d'image médicale par application de la matrice de cooccurrence » Présenté par : Mr : SIKEL Hakimet Melle : AMARA Karima« Université Mouloud Mammeri De Tizi-Ouzou ».
- [3]-Mémoire pour l'obtention de diplôme de master « Segmentation des IRM-3D en utilisant les champs aléatoires de Markov cachés et cuckoosearch technique. » Présenté par : Melle. Boukabene Randa et Mlle. Ouali Assia « Ecole nationale de l'informatique 2019/2020 ».
- [4]-Mémoire pour l'obtention de diplôme de master « Segmentation d'images IRM : application sur des IRM de patients atteints d'Alzheimer. » Présenté par : Mr HEBIB Mahdiet Mr SBAIHI Toufik« Université Mouloud Mammeri De Tizi-Ouzou ».
- [5]-Guerrou El-Hachemi. « Performances dans la Segmentation d'images médicales ». Thèse de doctorat. Oued-Smar-Alger : National School of Computer Science,juil. 2018.
- [6]-BOUCHAOUR Nadjat. « Approches collectives basées datamining etmachine-learning pourl'interprétationd'images et reconnaissance d'objets ». Thèse de doctorat. Université de 20 Août 1955 Skikda.
- [7]J.-P. COCQUEREZ et S. PHILIPP, « Analyse d'images : filtrage et segmentation », page3 -266, Paris, (février 2003).
- [8]-J.Gomes and O.Faugeras. Reconciling distance functions and level sets. INRIA Research Report, RR-3666, Avril 1999.
- [9]-C.Houassine, segmentation d'images par une approche biomimétique hybride. Université m'hamed bougara- boumerdes. 2012.
- [10]- M. Kass, A. Witkin And D. Terzopoulos, snakes : active contour modèles, international journal of computer vision, 1(4): pp321–331, 1987. cc
- [11]- M. Melliani, segmentation d'image par coopérationrégions-contours, magistère en informatique, école national supérieur d'informatique, 2012.
- [12]-SarraBenfriha et Asma Hamel-Segmentation d'image par Coopération régioncontours-Mémoire Master Professionnel Université KasdiMerbah-Ouargla-2015/2106.

[13]- 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 Cepaduès, pp. 193-198. 1999.

[14]- Yuan DONG –Thèse : Modélisation probabiliste de classifieurs d'ensemble pour des problèmes à deux classes- Université De Technologie De Troyes-2013.

[15]- 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.

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

Vaste Marge (SVM)-2007/2008.

[17] -M.Melliani, segmentation d'image par coopération régions-contours, magistère en informatique, école nationale supérieure d'informatique, 2012.

[18] - MAITRE, H. (2004). Le traitement des images.

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

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

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

[22] -Boucheham Bachir – cours 2eme année master Académique 'data Manning' – Université des Kikda 20 août 2016

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

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

[25] -C.Houassine, segmentation d'images par une approche biomimétique hybride. Université M'hamed Bougara-Boumerdes. 2012.

[26] - 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

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

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

[29] - 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.

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

[31]-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 MohamedBoudiaf USTO -BP 1505 El Mnaouer -ORAN – Algérie.

[32]- OumiloudHoriya, Mokeddem Asma, Classification non supervisée : Application de k-means. Diplôme de Licence en Informatique. Université Abou BakrBelkaid– Tlemcen,2014.

[33]-RawdhaBahri, NourelhoudaHadjadj, Segmentation d'Image par Classification (k-Means). Diplôme de Master en télécommunication. Université kasdi Merbah– Ouargla, 2020.

[34]-L. Houria, B. Oumessàd, « Classification Des Feuilles de Plantes à Base de Moment deHu », Mémoire Master Académique en informatique, 2014.

[35]- 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.~

[36] -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

[37] - B.youns, Tout ce que vous voulez savoir sur l'algorithme K-Means. 10avril2018.

[38] -A. Yassine. « Algorithme k-Moyennes » Cofé scientifique. PP : 5-6-18.

[39] -Hausmane Issarane, « Machine Learning. K-means: Définition& Avantages / Inconvénient ». 5 Mars 2019.

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

[41] -Geeksforgeeks:<https://www.geeksforgeeks.org/libraries-in-python/>, consulter le:08/04/2024

[42] - <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/> consulter le : 08/06/2024

[43] -<https://geekflare.com/fr/google-colab/> consulter le : 08/06/2024



Autorisation de Dépôt de Mémoire de Master

Je soussigné: *Mazouzi Smaïl*.....

Certifie que l'étudiant(e) : *Slimane Tich Tich HABIBA et Masbah Imes*

Spécialité : *Systèmes Informatique (S.I.)*.....

Ayant soutenu le projet intitulé : *Techniques d'apprentissage Automatique pour la Segmentation d'Images Médicales (I.R.M.)*.....

A apporté les corrections nécessaires sur son manuscrit de Master

Signature de l'encadreur



عدد: 07/18/2024

## بطاقة معلومات خاصة بمذكرة التخرج

رقم التسجيل :

اسم و لقب الطالب :

2194303566 • د. سليمان تيش تيش حبيبة

1802614631 • مهدياح تيبيايس

اسم و لقب المشرف على المذكرة : الأستاذة د. معز زوي سماح

عنوان المذكرة : Techniques d'apprentissage

Automatique pour la segmentation d'images  
Médicales (IRM).

القسم : الإعلام الإلكتروني

المستوى : ماجستير

التخصص : أنظمة المعلومات (S.I.)