MINISTRY OF HIGHER EDUCATION

AND SCIENTIFIC RESEARCH

UNIVERSITY KASDI MERBAH OUARGLA

Faculty of Applied Sciences

Department of Electronics and Telecommunications

Memory

ACADEMIC MASTER

Field: Science and Technology.

Branch : Electronic

Specialty: Electronics of embedded systems

Presented by:

**Derdouri Anis & Bakiri Yanis Abdelkarim**

**Theme:**

# Segmentation of medical images using the KFCM method

**Evaluation date : 21/06/2023**

**The jury composed of :**

| | | |
|---|---|---|
| Mr MANSEUR Abdelghani | (UKM Ouargla). | President |
| Mr NASRI Nadjib | MAA (UKM Ouargla). | Supervisor |

**Academic year : 2022/2023**

# Special thanks

Thanks to our Lord "ALLAH" who has blessed us with The wonderful faculty. Praise be to our Creator who has encouraged us to To acquire knowledge.

At the end of this work, we take this opportunity To address our thanks to all our teachers of the university KASDI MERBAH Ouargla who have contributed to our training.

We express our deep gratitude to our supervisor NASRI NADJIB, it was a great pride and an honor for us to work under the guidance of a man for whom we have great admiration, and we thank him for his help and his precious advice.

Our thanks also go to the jury for their interest to this work.

Finally we thank all those who have contributed in any way to make this work.

# Index

# List of figures

## Chapter I :General information about image processing



## Chapter II: Segmentation By Unsupervised Methods



### Chapter III:Results and discussion

# List of tables

### Chapter II: Segmentation By Unsupervised Methods



### Chapter III:Results and discussion

# General introduction

# General introduction

Medical imaging technologies, such as magnetic resonance imaging (MRI), have revolutionized healthcare by providing detailed visual representations of internal structures.

However, the amount of data generated by medical images poses significant challenges for accurate and efficient interpretation. Image segmentation techniques offer a solution by partitioning the images into meaningful regions, facilitating targeted analysis and diagnosis.

The primary objective of this work is to develop and evaluate image segmentation algorithms, namely k-means[11], fuzzy c-means[5], and KFCM[6], for medical images, particularly focusing on MRI data. The aim is to provide automated and reliable segmentation methods that can assist doctors in their diagnostic processes and future medical interventions.

Our work was divided into three parts, in the first chapter; we began by providing a general reminder about images and the image processing process. This introductory section likely aimed to familiarize the reader with fundamental concepts related to images and the steps involved in image processing

In the second chapter, we will explore the topic of segmentation by unsupervised methods, emphasizing the importance of accurate and efficient segmentation for various medical tasks. This chapter set the foundation for exploring the KFCM method in the subsequent chapter.

In the third chapter, a comprehensive comparison was conducted between the K-Means, Fuzzy C-Means, and Kernel Fuzzy C-Means algorithm for segmenting medical images. The evaluation was performed using metrics such as Jaccard coefficient[10], partition coefficient[15], and Dice coefficient [15] to assess their segmentation accuracy and robustness.

The completion of this work is expected to contribute to the advancement of medical image segmentation techniques, specifically in the context of MRI data. The proposed algorithms have the potential to improve segmentation accuracy and efficiency, ultimately facilitating easier analysis and interpretation of medical images by medical professionals.

# Chapter I :General information about image processing

# I. Introduction :

Image processing is a rapidly growing field that deals with the manipulation of digital images using computer algorithms. It is a multidisciplinary field that draws upon knowledge from mathematics, computer science, and electrical engineering, among others. Image processing has a wide range of applications,The first chapter of this thesis aims to provide a general overview of image processing. And introduction to the basic concepts and terminology used in image processing

# II. Definition of an Image :

The definition of an image is a visual representation of a person, object, or event, created through various mediums such as painting, sculpture, drawing, photography, film, etc. It can also be considered a structured set of information that, after being displayed on a screen or other display device, has meaning for the human eye to interpret the visual representation.[2]

### II.1 Grayscale image:

Simple images have an intensity value based on a defined number of grayscale. Black represents the value 0 and white represents the value 255 in the grayscale levels palette [1]

### II.2 Black and white image :

A black and white image has elements with a value of either "1" or "0"." [1]

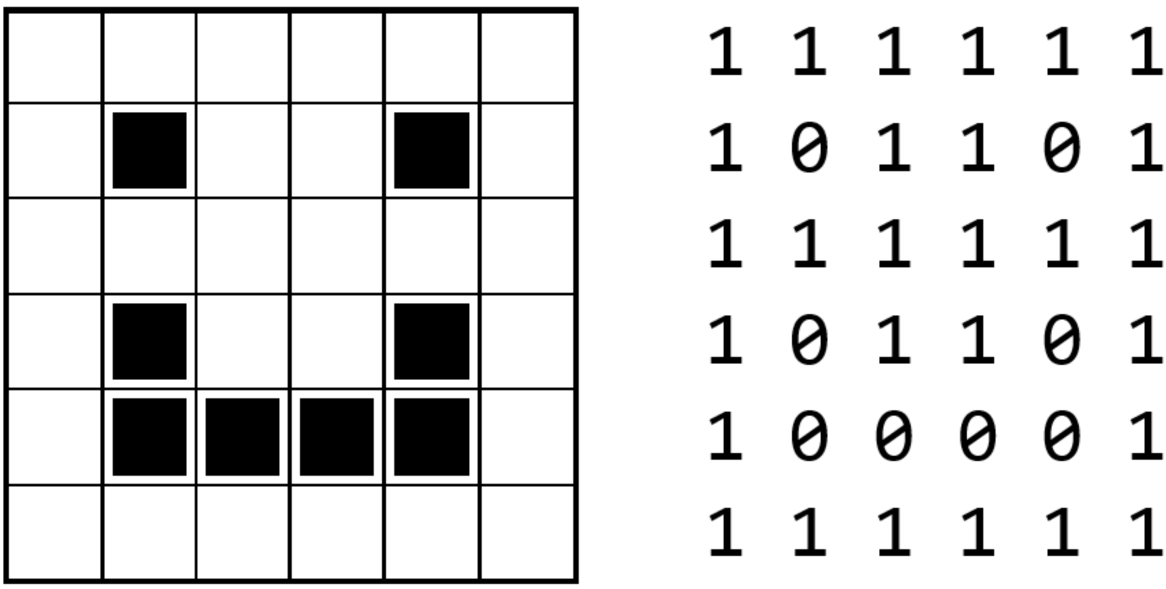



Figure I. 1: Black And White image

### II.3 Digital Images:

A digital image is a representation of a real image as a set of numbers that can be stored and handled by a Informatic tool. In order to translate the image into numbers, it is divided into small areas called pixels (picture elements). For each pixel, the imaging device records a number, or a small set of numbers, that describe some property of this pixel, such as its brightness (the intensity of the light) or its color. The numbers are arranged in matrix that correspond to the vertical and horizontal positions of the pixels in the image.[2]

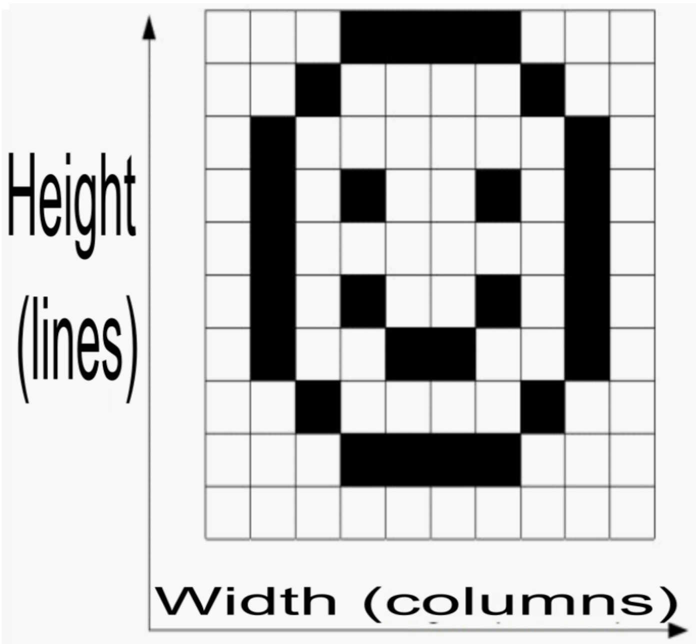



Figure I. 2: Digital representation of an image

## III. The characteristics of a digital image:

### III.1 The Pixel:

The pixel is the smallest point of an image, it is a calculable entity that can receive structure and quantification. If the bit is the smallest unit of information that a computer can process, the pixel is the smallest element that display, or printing hardware and software can manipulate. The image can be displayed as a group of pixels as shown in the figure below: [1][19]

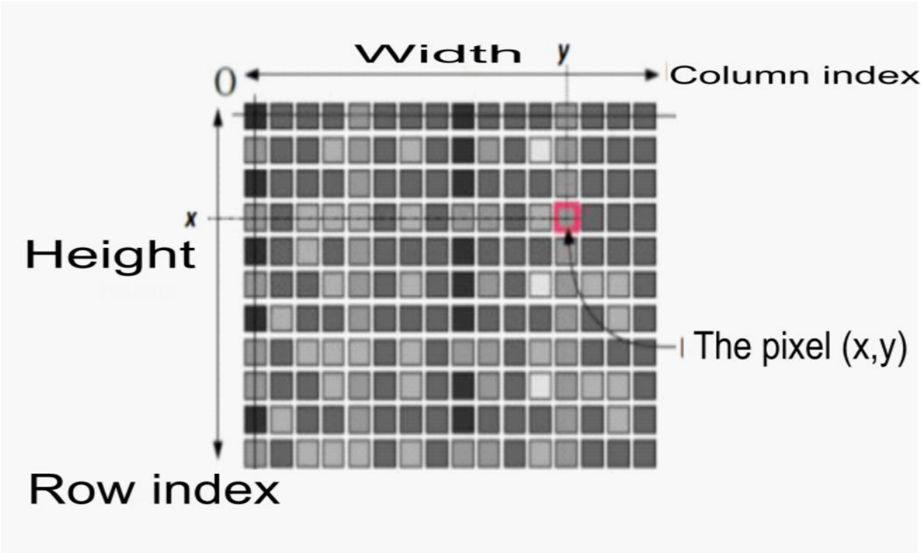



Figure I. 3: Representation of the Pixels

### III.1.1 Representation of pixels in an image :

A digital image is represented by a matrix of dimension (m x n) of hardware elements that are pixels. The value of each pixel represents a color considered in the image.

### III.2 The size of an image :

The Image Size consists of the width and height of an image. It can also be understood as the image resolution. Usually, these values are given in pixels and the following format 1024x981 but it can also be given in : inches (Ex : 24 x 36 inches) or in dpi (Ex: 300dpi) [1]

### III.3 Resolution :

Image resolution refers to the level of detail and clarity in an image, usually measured in pixels per inch (PPI) or dots per inch (DPI). It describes the number of pixels or dots that make up an image and determines the sharpness and quality of the image. The higher the resolution, the more pixels or dots there are in the image, resulting in a clearer and more detailed image. Image resolution is an important factor to consider when creating or editing images, as it affects how the image will appear when displayed or printed.


Figure I. 4: 1920x1280 ( Original Image)

(a) (b)

Figure I. 5: (a) 120*80, (b) 200*133

### III.4 The Noise :

Noise in an image is considered a phenomenon of sudden variation in the intensity of a pixel in relation to its neighbors. It comes from the illumination of the optical and electronic devices of the sensor.[1]

### III.4.1 Sources of Noise :

- **Sensor Used:** Poor quality of the sensor is one of the main causes of degradation. [2]
- **Acquisition Context**: Blurring, motion, or changes in lighting conditions are often related to things that can modify the image. [2]
- **Nature of the Scene:** The scene itself can be a source of disturbance, for example: weather conditions, reflections, dust particles [2]
- **Sampling:** Sampling can cause degradation in the quality of the digital image.
- **Electromagnetic interference (EMI):** it is caused by the presence of electromagnetic fields that interfere with the signal being transmitted or received.
- **Quantization noise**: it is caused by the rounding errors that occur when converting an analog signal to a digital signal. [2]
- **Crosstalk noise:** it is caused by signals from one channel interfering with signals from another channel in a communication system.
- **Instrumental noise:**It is a noise commonly encountered in measuring systems, which is due to the electronics present in the instrumentation.
- **Experimental noise**: Experimental noise refers to any reason that can modify the intensity of light received by the detector. These reasons include ambient light, movement of the subject between the source and the sensor, or motion artifacts. It refers to a sudden modification of the intensity returned by a sensor without any possible physiological cause, caused by the movement or motion of the subject.
- **Physiological noise:** It represents the physiological elements that do not provide any information about brain activity, such as heartbeats or movements during breathing. It does not prevent the interpretation of data, but it needs to be taken into account to understand the acquired signals.

- **Motion artifact:** Motion artifact noise in refers to the degradation of image quality caused by unwanted motion or movement during image acquisition or by the surrendings .The noise caused by motion artifact can be particularly problematic in MRI because the images are sensitive to even small movements, which can result in blurred, distorted, or unusable images.

- **Digital noise:** Digital noise refers to the unwanted variation or random fluctuations in a digital signal or image. It is a common problem in digital imaging and can result in degraded image quality, reduced contrast, and reduced dynamic range.

## III. 5. Type of noises:

### III. 5.1. Gaussian Noise:

This is a type of random noise that is characterized by a normal or Gaussian distribution of signal fluctuations. It can be caused by various factors, such as electrical interference, signal amplification, or quantization errors. [3]

(a) (b)

Figure I. 6: (a) Original Image ( grayscale), (b) Gaussian Noise

### III.5.2 Salt and Pepper Noise:

This type of noise appears as randomly occurring black and white pixels in an image. It can be caused by various factors such as faulty pixels in a camera sensor, signal transmission errors, or errors in the analog-to-digital conversion process. [3]

(a) (b)

Figure I. 7: (a) Original Image ( grayscale), (b) Salt and Pepper Noise

### III.5.3 Uniform noise:

is a type of digital noise that appears as a constant signal variation with a fixed amplitude. It is characterized by a constant distribution of signal fluctuations and is typically caused by various factors such as electrical interference, data transmission errors, or faults in digital sensors.

(a) (b)

Figure I. 8: (a) Original Image ( grayscale), (b) *Uniform noise*

### III.6. Noise measurement:

The Signal-to-Noise Ratio (SNR) is used to estimate the quality of image i2 compared to a reference image i1. [2]

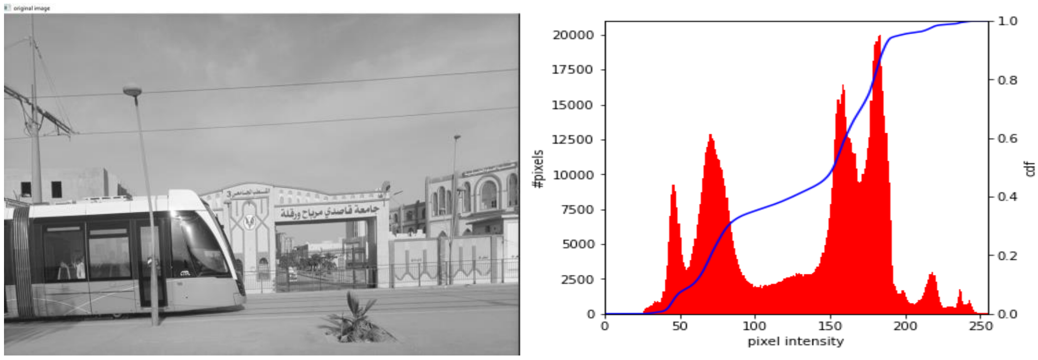
$$\mathrm{SNR}(\mathrm{I1}/\mathrm{I2}) = 10.\log 10 \left[\frac{\sigma^2(\mathrm{I1})}{\sigma^2(\mathrm{I1}-\mathrm{I2})}\right] \qquad \mathrm{I.1}$$

{If SNR $\geq 20$ Then I2's quality is good

{If SNR $\leq 10$ Then I2's quality is bad

### III.7 Histogram:

In an image, a histogram is a graphical representation of the distribution of pixel values. It displays the number of pixels at each brightness level or color intensity in the image. The horizontal axis of the histogram represents the range of pixel values, usually from 0 to 255 for an 8-bit grayscale image or for each color channel in a color image. The vertical axis represents the number of pixels that have the corresponding brightness level or color intensity. [1][19]




(a) (b)

Figure I. 9: (a) Original Image ( grayscale), (b) *Histogram*

### III.8 Luminance:

it is the degree of brightness of the points in the image. It is used instead of the word "brilliance", which corresponds to the brightness of an object [1]

(a) (b)

Figure I. 10: (a) Original Image, (b) *Original image's brightness enhanced at 50%*

### III.9 Contrast:

The contrast of an image is an intrinsic property that quantifies the difference in brightness between its light and dark areas. Contrast characterizes the luminous distribution of an image. Visually, this can be considered as a spread of the brightness histogram of the image [1][19]

(a) (b)

Figure I. 11: (a) Original Image, (b) *Original image's contrast adjusted at 50%*

## IV. Image processing :

Image processing has many steps or stages that are used to process an image. it starts with inputting the image into a computer system and ends with outputting a processed image or image features that can be used for further analysis or applications. This process may involve various techniques such as image filtering, image segmentation, feature extraction, and classification.

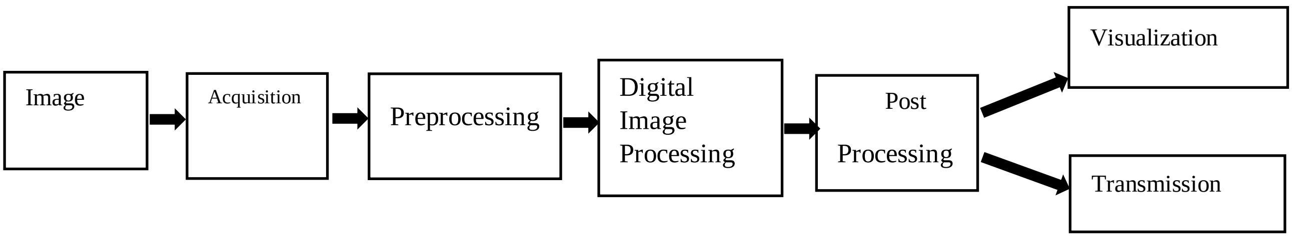



Figure I. 12: Image processing diagram.

The general image processing consists of the following stages:

**Acquisition:** This is the stage where the image is captured or obtained from a source. The image can be obtained from a camera, scanner, or any other imaging device.

- **Preprocessing:** This stage involves pre-processing the acquired image to improve its quality, such as correcting for distortion, adjusting the contrast, and removing noise.
- **Segmentation:** In this stage, the image is divided into meaningful parts or regions. This stage is crucial for object recognition and tracking. Techniques used in this stage include thresholding, edge detection, and region growing.
- **Feature extraction:** This stage involves extracting specific features from the segmented image regions. The extracted features can be used to represent the image in a way that is more useful for analysis or classification.
- **Classification**: This stage involves classifying the image based on the extracted features. This stage is important in applications such as object recognition, face recognition, and medical image analysis.
- **Post-processing:** This stage involves further processing of the classified image, such as refining the segmentation results or applying a filter to reduce noise.

## IV.1. Image processing applications :

Image processing has numerous applications in a variety of fields. Here are some of the most common:

### IV.1.1. Medical Imaging:

Image processing is used in medical imaging to enhance images, detect abnormalities, and aid in diagnosis. This includes X-rays, CT scans, MRI scans, and ultrasound images. [16]

### IV.1.2. Surveillance:

Image processing is used in video surveillance to detect suspicious behavior or events in real-time. This is done by analyzing the video feed and alerting security personnel to potential threats. [16]

### IV.1.3. Robotics:

Image processing is used in robotics to enable machines to "see" and interpret the physical world. This includes autonomous vehicles, drones, and robots used in manufacturing and other industries. [16]

### IV.1.4. Entertainment:

Image processing is used in the entertainment industry to create special effects, enhance images, and improve the quality of films, TV shows, and video games[16].

### IV.1.5. Remote Sensing:

Image processing is used in remote sensing applications to analyze satellite and aerial images of the Earth's surface. This includes monitoring land use, mapping changes in vegetation, and identifying areas of environmental concern. [16]

### IV.1.6. Biometrics:

Image processing is used in biometric applications, such as facial recognition and fingerprint scanning, to identify individuals and enhance security.

### IV.1.7. Quality Control:

Image processing is used in quality control applications to inspect products and detect defects in manufacturing processes.

## V. Medical imaging:

is a non-invasive tool used in medicine for diagnostic and treatment purposes, allowing doctors to recreate images of different body parts. It is a central component of modern medicine, enabling doctors to diagnose injuries and diseases without being intrusive.[20]

### V.1. Types of medical imaging procedures :

a) **X-rays** : X-rays use ionizing radiation to produce images of a person's internal structure by sending beams through the body. These are absorbed at different levels depending on the density of the tissue. X-ray radiation can generate three kinds of medical images; conventional X-ray imaging, angiography and fluoroscopy.[20]


Figure I. 13: X-ray Radiography

b) **Computed Tomography (CT) Scan:** Also commonly referred to as a CT scan, Computed Tomography is an imaging technique that combines multiple X-ray images taken from different angles. This produces detailed cross-sectional internal images.[20]

Figure I. 14: Computed Tomography Scanr (CT Scan)

c) **Magnetic Resonance Imaging (MRI):** Magnetic Resonance Imaging (MRI) is a technology that uses radio waves and a magnetic field to provide detailed images of organs and tissues. The type of radiation in this kind of imaging technique generates images of the soft tissues, omitting the bones.[20]

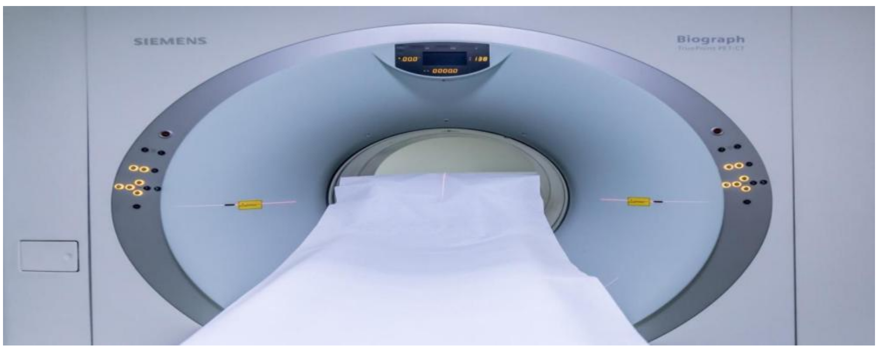



Figure I. 15: Magnetic Resonance Imaging (MRI) Scan

**d) Ultrasound imaging:** also known as medical sonography or ultrasonography, it uses high frequency sound waves to create images of the inside of the body. The ultrasound machine sends sound waves into the body and is able to convert the returning sound, echoes, into a picture.[20]

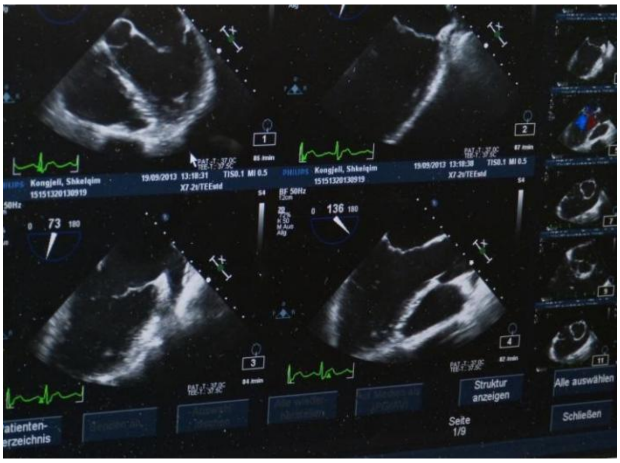



Figure I. 16: Ultrasound Echography

## VI. Conclusion:

This chapter covered some generalities about image processing, including the definition and characteristics of digital images, the common types of noise that can affect images, and the basic steps of image processing . We also briefly discussed medical imaging. In the next chapter, we will discuss image segmentation, which is a fundamental step in many image-processing applications.

# Chapter II: Segmentation By Unsupervised Methods

## I. Introduction:

Image segmentation is the process of dividing an image into multiple segments or regions; this can be useful for a wide range of applications, such as object recognition, tracking, and image compression.

Segmentation allows us to extract useful information from an image, by breaking it down into smaller; this information can be used for a variety of purposes, such as identifying different objects in the image, measuring their size, shape and color.

In many applications, segmentation is used to extract relevant features from an image, which can be used for further analysis. For example, in medical imaging, segmentation is often used to identify and isolate specific organs or tissues in an image,

Overall, image segmentation is an important and widely used technique in computer vision and image processing, with many practical applications in fields such as medicine, robotics, surveillance, and more there are many popular segmentation algorithms but in this chapter we will focus on K-means clustering and Fuzzy C-means clustering.

## II. Definition of segmentation:

The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. Segmentation can be performed at various levels of detail, ranging from a coarse level where the image is divided into just a few segments, to a fine level where each individual pixel is assigned a unique label or segment. The choice of segmentation level depends on the specific task and the type of image being analysed

Segmentation can be performed manually by human experts or using automated algorithms that are designed to identify patterns or features in the image data. Common techniques for automated segmentation include thresholding, clustering …..

## III. Classification segmentation:

The classification segmentation approach can be **supervised** or **unsupervised**.

### III.1. Supervised methods:

Supervised methods are machine learning techniques in which a model is trained on a previously labelled data set. The objective of supervised learning is to predict an output (or class) from a given input.

In the context of image segmentation, supervised methods can be used to classify pixels or regions of an image into different classes or categories. To do this, a set of previously labelled images is used to train a classification algorithm, such as a neural network, SVM (support vector machine) or regression.

Supervised methods have the advantage of providing accurate and consistent results, as the model is trained on a labelled data set. However, they require an appropriately labelled training data set, which can be costly and time-consuming. In addition, they can be prone to overfitting, where the model is too complex and fits the training data too well, which can lead to poor generalisation to the test data.

### III.2. Unsupervised methods:

Unsupervised methods are machine learning techniques in which a model is trained on unlabelled data. Unlike supervised methods, unsupervised methods do not require prior labels

In the context of image segmentation, unsupervised methods can be used to group pixels or image regions according to their common characteristics. For this purpose, clustering algorithms such as k-means, FCM (fuzzy C-means), DBSCAN (Density-Based Spatial Clustering of Applications with Noise) or hierarchical clustering can be used.

Unsupervised methods have the advantage that they do not require manual labelling of the data, which can be tedious and expensive. In addition, they can be used to discover hidden patterns in the data and identify unexpected groups or categories. However, they can be subject to variability in results, as results can vary depending on the input parameters of the algorithm and the data itself.

## IV. K-means Algorithm:

Several researchers in different fields first proposed the K-means algorithm independently, Its development is due to different authors. Some of the key contributors to the development of the k-means algorithm include:

**Hugo Steinhaus**: The k-means algorithm has been attributed to the Polish mathematician Hugo Steinhaus, who proposed a similar clustering method in the 1950s called the "minimum variance method". [11][12]

**Stuart Lloyd**: The k-means algorithm is also attributed to Stuart Lloyd, an electrical engineer at Bell Labs who proposed the algorithm in his 1957 paper "Least Squares Quantization in PCM". [11][12]

**John MacQueen**: The k-means algorithm is also credited to John MacQueen, a statistician who independently proposed a similar clustering method in 1967, which he called "k-means clustering". [11][12]

**Jancey**: Jancey, from the Department of Botany, School of Biological Sciences, University of Sydney, , is also credited with in one of his articles in 1966 titled “Multidimensional Group Analysis” [11][12]

The k-means algorithm is a clustering algorithm that aims to partition a set of data points into k clusters based on their similarity. The algorithm works by iteratively assigning each data point to the nearest cluster centroid and then updating the centroid based on the mean of the data points in the cluster. This process is repeated until the clusters converge and no more reassignments occur. [8]

### IV.1. Main function

The main function of the K-means algorithm is to partition a given set of data points into k clusters based on their similarity. This is achieved by iteratively minimizing the sum of squared distances between each data point and its assigned cluster centroid.

The k-means algorithm is commonly used in data mining, machine learning, and image segmentation, among other applications. It is particularly useful in situations where the data set is large and the number of clusters is not known in advance.

The main steps of the k-means algorithm are: [4]

1. Random selection of the initial position of the K clusters.
2. Assign the objects to a cluster according to a distance minimization criterion (Usually according to a Euclidean distance measure).
3. Once all objects are placed, recompute the K centroids.
4. Repeat steps 2 and 3 until no more reassignments are made

Table II. 1: kmeans algorithm

**The Convergence**: The algorithm converges when the cluster assignments no longer change or when a maximum number of iterations is reached.

Overall, the k-means algorithm is a powerful tool for data clustering and has a wide range of applications in various fields**.**

### IV.2 Diagram of the kmeans algorithm:

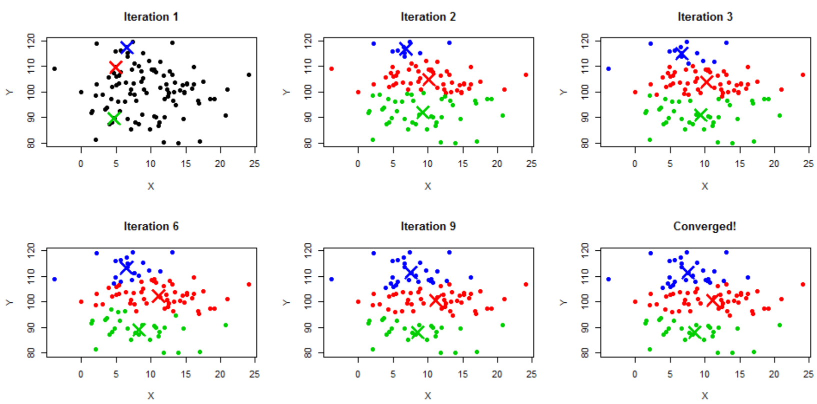



Figure II. 1:Diagram explaining how the kmeans works

### IV.3. Algorithm

We will present K means-clustering algorithm and the aim is to minimize the sum of squared distances between each data point and its assigned cluster centroid function defined as follow: [8]

$$\boldsymbol{j} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} ||\boldsymbol{x}_{\boldsymbol{j}}^{(\boldsymbol{j})} - \boldsymbol{c}_{\boldsymbol{j}}||^2 \qquad \textbf{(II.1)}$$

**Where:**

- $||x_j^{(j)} - c_j||^2$ is a chosen distance measure between a data point $x_j^{(j)}$ and the cluster center $c_j$ , is an indicator of the distance of the n data points from their respective cluster centres.
- $k$ is the number of clusters
- $n$ number of cases
- $\boldsymbol{c_j}$ Centroid for cluster j
- j the objective function

## V. FCM Algorithm:

Dunn first proposed the Fuzzy C-Means (FCM) algorithm in 1973 as an extension of the classic K-Means algorithm, which is a hard clustering algorithm that assigns each data point to a single cluster. In FCM, each data point is assigned a degree of membership to each cluster, which allows for soft clustering and overlapping clusters.

The FCM algorithm was further developed and popularized by Bezdek in the late 1970s and early 1980s. Bezdek introduced the concept of fuzzification, which allows the degree of membership of a data point to a cluster to range between 0 and 1, rather than being restricted to binary values of 0 or 1 as in K-Means. This concept of fuzzification was a major contribution to the field of clustering and fuzzy logic.

Since then, the FCM algorithm has been extensively studied and applied in many different fields, including image processing, data mining, and pattern recognition.

The algorithm has also been extended and modified in various ways, such as incorporating constraints, incorporating prior knowledge, and handling missing data.

The Fuzzy C-Means (FCM) algorithm is used to partition a dataset into several groups or classes based on certain common characteristics.

This algorithm has multiple functions such as image segmentation and recognition and for data analyse.

FCM is derivative from K means, which means it is an upgrade for K means when the k is the number of clusters, and the user generally writes it. [5]

### V.1. The main function of the Fuzzy C-Means Algorithm

The main function of the Fuzzy C-Means (FCM) algorithm is to partition a given dataset into a set of clusters or groups. FCM is a soft clustering algorithm, which means that each data point is assigned a degree of membership for each cluster rather than being strictly assigned to a single cluster. The degree of membership of a data point for a given cluster represents the degree of similarity or belongingness of the data point to that cluster.

The FCM algorithm works by iteratively updating the centroid vectors of the clusters and the degree of membership of each data point to each cluster. The algorithm minimizes an objective function that measures the total distance between each data point and the centroid vector of its assigned cluster, weighted by the degree of membership of the data point to the cluster.

### V.2. Membership degree:

Membership degrees, also known as membership values or membership degrees of fuzzy sets, refer to the degree of membership or belongingness of an element to a particular set in fuzzy logic.

Membership degrees are represented by values between 0 and 1, where 0 indicates no membership or complete non-membership, and 1 represents full membership or complete inclusion in the set. Intermediate values between 0 and 1 indicate degrees of partial membership, reflecting the extent to which an element belongs to the set.

### V.3.Diagram for Fuzzy c means algorithm :

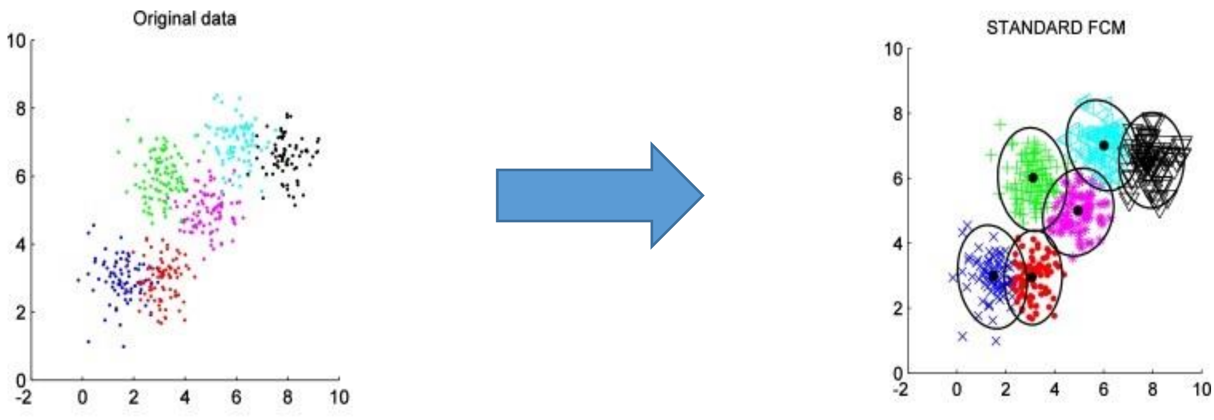



Figure II. 2 :Diagram explaining how the fuzzy c means works

### V.4. Algorithm

We will present the fuzzy C means-clustering algorithm, which is very similar to the k means algorithm, and the aim is to minimize the objective function defined as follow: [16]

$$Jfcm=\sum_{j\in\Omega}\sum_{k=1}^{c} u_{jk}^{m} \ || y_j - v_k ||^2 \qquad \textbf{(II.2)}$$

Where :

- j is an index representing a data point in the dataset, where j ∈ Ω denotes that j belongs to the set of data points Ω.
- k is an index representing a cluster center, ranging from 1 to c (the total number of clusters).
- $\mathbf{u_{jk}}$ represents the membership value of data point j in cluster k. The membership value indicates the degree to which the data point belongs to a particular cluster.
- m is a fuzziness parameter (usually greater than 1) that controls the degree of fuzziness in the clustering.
- $\mathbf{y_j}$ represents the feature vector of data point j.
- $\boldsymbol{v_k}$ represents the feature vector of cluster center k.
- $|| \mathbf{y_j} - \boldsymbol{v_k} ||^2$ is the squared Euclidean distance between data point j and cluster center k.

The distance between points and centres:

$$D_{ik} = \sqrt{(X_k - X_i)^2} \qquad \textbf{(II.3)}$$

The convergence:

$$\text{Max}_{ik} |U_{ik}(1) - U_{ik}(0)| <= \text{error} \qquad \textbf{(II.4)}$$

Step 1) initialization of the number of clusters,m,and t (maximum number of iterations ) and Obtain the initial class centers

Step 2: Calculate the degrees of membership using the following formula:

$$u_{ij} = \frac{1}{\Sigma_{k=1}^{c} \left(\frac{||y_j - v_i||}{||y_j - v_k||}\right)^{\frac{2}{m-1}}} \tag{II. 5}$$

Step3: Update the class centers using the formula:

$$v_k = \frac{\Sigma_{j\in\Omega}\, u_{jk}^{q} \cdot y_j}{\Sigma_{j\in\Omega}\, u_{jk}^{q}} \tag{II. 6}$$

Step 4: Repeat steps 2 and 3 until convergence is reached:

$$\| j^{new} - j^{old} \| < \varepsilon \tag{II.7}$$

Table II. 2: FCM algorithm

### V.3. Distances:

**a) Euclidean distance**: the Euclidean distance is the straight length of a line connecting the two points. [21]

$$dEucl(p,q) = \sqrt{(X2 - X_1)^{2} + (Y_2 - Y_1)^{2}} \tag{II.8}$$

**b) The Tchebychev distance:**

The Tchebychev distance between two vectors or dimensional points n is the maximum absolute magnitude of the differences between the coordinates of the points. The distance of tchebychev evaluates the maximum absolute value of the differences between the coordinates of a pair of objects. [21]

$$d_{cheb}(p,q) = \max |p_i - q_i| \tag{II.9}$$

**c) The Manhattan Distance:**

It is suitable for measuring theoretical distances; it allows to calculate the distance between two data points on a uniform grid. [21]

$$d_{Manh}(p,q) = |x_{2} - x_1| + |y_2 - y_1| \tag{II.10}$$

**d) The Minkowski Distance:**

Minkowski Distance is a distance between two points in dimensional space n. It is a generalization of the Euclidean, Manhattan and Chebychev distances. [21]

$$d_{Mink}(q,p) = \sqrt[\lambda]{\sum_{i=1}^{n} (|\ q_i - p_i\ |)^{\lambda}} \qquad \textbf{(II.11)}$$

Where λ is the order of the Minkowski metric. For different values of λ, the distance can be calculated in three different ways:

$\lambda = 1 \rightarrow$ Manhattan Distance (métrique $L^1$)

$\lambda = 2 \rightarrow$ Euclidienne Distance (métrique $L^2$)

$\lambda = \infty \rightarrow$ Tchebychev Distance (métrique $L_\infty$)

#### e) Mahalanobis distance:

The distance of Mahalanobis allows to calculate the distance between two points in a space with p dimensions, taking into account the structure of variance-covariance on these p dimensions [21]

$$dM^2 = (x1 - x2) \sum{-1} (x1 - x2) \qquad \textbf{(II.12)}$$

#### f) The distance Gaussian :

the distance Gaussian (also known as a Gaussian kernel or Gaussian function) is a mathematical function used in various contexts, including clustering and smoothing. It defines a probability distribution that is often used to calculate the similarity or dissimilarity between data points based on their distances By using kernel function, distance between $\Phi(x)$ and $\phi(y)$ defined as :

$$\boldsymbol{d^2(\Phi(x), \phi(y)) = k(x,x) - 2k(x,y) + k(yy)} \qquad \textbf{II.13}$$

## VI. Kernel Fuzzy C Means Algorithm：

Kernel Fuzzy C-Means (KFCM) is a variation of the Fuzzy C-Means (FCM) algorithm that incorporates a kernel function to transform the original data space into a higher dimensional space, where the clusters are better separated. KFCM is particularly useful in cases where the data is non-linearly separable and cannot be well clustered in the original data space.

In KFCM, the kernel function maps the original data space into a higher dimensional feature space, where the FCM algorithm is then applied to perform soft clustering. The kernel function is chosen based on the characteristics of the data, and commonly used kernel functions include Gaussian, polynomial, and sigmoid kernels.

The KFCM algorithm iteratively updates the centroids and membership degrees of each data point until convergence is achieved. The degree of membership of a data point to a cluster is calculated based on the distance between the data point and the centroid of the cluster in the kernel feature space.

### VI.1. History of Kernel fuzzy c means

The use of kernel functions in clustering algorithms dates back to the 1960s, but it was not until the late 1990s and early 2000s that the idea of combining kernel functions with fuzzy clustering algorithms gained popularity. In 1999, Zhang and Sugeno introduced the concept of kernel-based fuzzy clustering, which used a kernel function to map the data into a high-dimensional feature space, where the clustering was performed.

KFCM was first introduced by Huang et al. in 1998, as a generalization of the FCM algorithm that incorporates a kernel function to transform the data into a high-dimensional feature space. Since then, KFCM has been extensively studied and applied in many different fields, including image processing, data mining, and pattern recognition.

Over the years, various modifications and extensions of KFCM have been proposed to improve its performance and applicability to different types of data. For example, some researchers have incorporated constraints or prior knowledge into KFCM, while others have used adaptive kernel functions that are tailored to the characteristics of the data.

### VI.2. The main function of Kernel Fuzzy C-Means

KFCM works by first applying a kernel function to map the original data space into a high-dimensional feature space, where the data points are better separated. The FCM algorithm is then applied to perform soft clustering of the data points in the feature space. The degree of membership of a data point to each cluster is calculated based on the distance between the data point and the centroid of the cluster in the feature space.

The KFCM algorithm iteratively updates the centroids and membership degrees of each data point until convergence is achieved. The resulting clusters can then be used to identify similarities or differences among the data points and to discover hidden patterns or structures in the data.

Overall, the main function of KFCM is to perform soft clustering of non-linearly separable datasets in a high-dimensional feature space, using a kernel function to capture the non-linear relationships among the data points.

### VI.3. Diagram for Kernel Fuzzy c means algorithm :

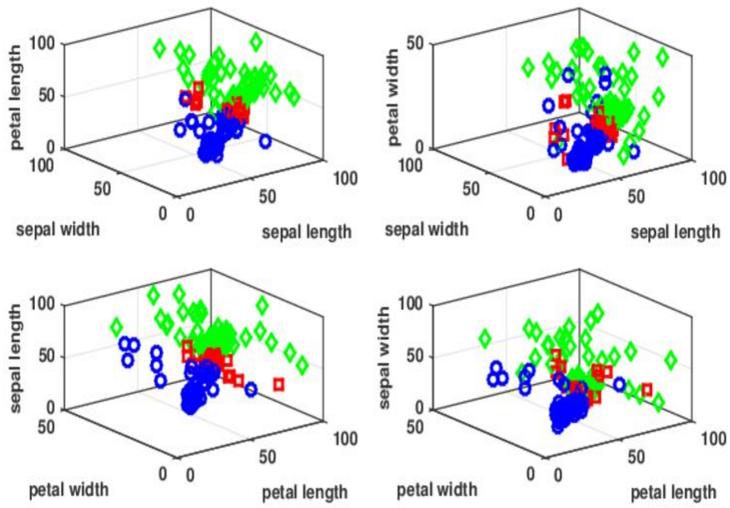



Figure II. 3 :Diagram explaining how the kernel fuzzy c means works

### VI.4. Algorithm:

We will present the kernel fuzzy C means-clustering algorithm, which is very similar to the FCM algorithm.

By combining Fuzzy C-Means Clustering and Kernel Function, Kernel Based Fuzzy C-Means Clustering was obtained with the mathematical model as follows:[6][7]

1)Input: c , tmax , m > 1 and ε > 0

2) Initialize the memberships $u_{ik}^{0}$.

3) For t =1,2,…, tmax , do:

(a) Update all prototypes $V_i^t$ with :

$$V_i = \frac{\sum_{k=1}^{n} C_k}{\sum_{k=1}^{N} u_{ik}^{m} K(x_k, v_i)} \quad \textbf{(II.14)}$$

(b) Update all memberships $u_{ik}^{t}$ with :

$$u_{ik} = \frac{(1-K(x_k, v_i))^{-1/(m-1)}}{\sum_{j=1}^{c} (1-K(x_k, v_i))^{-1/(m-1)}} \quad \textbf{(II.15)}$$

(c) Compute $E^t = \max_{i,k}|u_{ik}^{t} - u_{ik}^{t-1}|$ , if $E^t \leq \varepsilon$ , stop;

End;

Table II. 3: KFCM algorithm

## VII. Conclusion :

we have seen in this chapter three of unsupervised methods of clustering with their history the famous one is the Kmeans and its blurred version the Fuzzy c means and the last one its the enhanced version of FCM called kernel fuzzy c means .

In the next chapter, the tests on synthetic and real images will be presented and discussed.

# Chapter III:Results and discussion

# I. Introduction:

We dedicate this chapter to the presentation of the results, resulting from the application of the algorithms seen in the previous chapter on synthetic and real images.

# II. Synthetic image :

In order to compare the different algorithms, in particular their robustness to noise. We will be applying them to a synthetic image of dimension 128x128, that has two gray levels 0.7 and 0.3 respectively; with an additive Gaussian noise localized in the center in order to simulate these inhomogeneities.

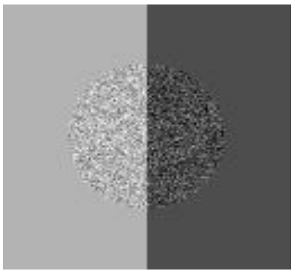


Figure III. 1: Synthetic image

# III. MRI image:

We will be also using an mri image with 181 pixels width and 217 pixels height.

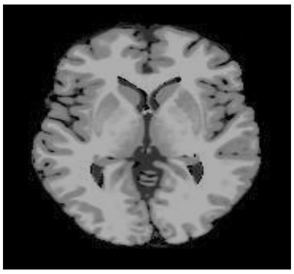
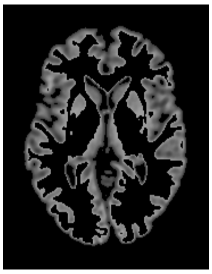


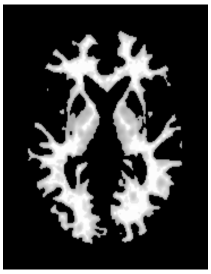
Figure III. 2: a real picture of the brain

The images used in this work are from the Brainweb database which is an online portal for the BrainWeb project, which is a collaborative effort between the McConnell Brain Imaging Centre (BIC) at McGill University and other research institutions. This project aims to provide a publicly accessible database of simulated brain images and related data for researchers working in the field of medical imaging and neuroscience.

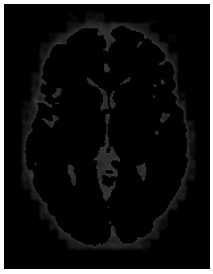
**III.1. Results of the three classes after segmentation:**



Gray matter



white matter



cerebrospinal fluid

Figure III. 3:the three classes of the brain after segmentation

## IV. Evaluation of image segmentation:

Segmentation is an essential step in image processing in the extent to which it conditions the interpretation that will be made on this footage. Many algorithms have been proposed during the last decades . They are based on different Approaches: contour, region, texture. Faced with the multitude of methods proposed, the problem of assessing the quality of segmentation becomes paramount...

Image segmentation evaluation is important for evaluating the performance of segmentation algorithms and for comparing different segmentation methods.

To evaluate the segmentation of an image, different evaluation measures can be commonly used include:

**IV.1. The Jaccard index**

The Jaccard similarity index (or Intersection over Union, IoU): This is a measure that compares the intersection and union between predicted segmentation and ground truth. He measures the extent to which the predicted segmentation is superimposed on the ground truth. A high IoU indicates precise segmentation.

The Jaccard index, also known as the Jaccard similarity coefficient, is a statistic used for gauging the similarity and diversity of sample sets. It was developed by Grove Karl Gilbert in 1884 .It was later developed independently by Paul Jaccard, and independently formulated again by T. Tanimoto. Thus, the Tanimoto index or Tanimoto coefficient are also used in some fields. However, they are identical in generally taking the ratio of Intersection over Union. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union the resulting value ranges from 0 to 1, where

1 indicates a perfect segmentation that precisely matches the ground truth. Higher values of IoU indicate better segmentation quality. [10]

$$jaccard\ index = \frac{Tp}{Tp+Fp+Fn} \tag{III.1}$$

Where TP are the true positives, FP the false positives and FN the false negatives.

**IV.2.The Dice coefficient**

The Sorensen–Dice coefficient is a statistic used to gauge the similarity of two samples. The botanists Thorvald Sorensen and Lee Raymond Dice who published in 1948 and 1945 respectively independently developed it.

The Dice coefficient, also known as the Dice measure or F1-weighted score, is a commonly used measure to assess the quality of image segmentation, especially in the context of binary segmentation.

The Dice coefficient is calculated from the similarity between the predicted segmentation and the ground truth When applied to Boolean data, using the definition of true positive (TP), false positive (FP), and false negative (FN), it can be written as: [15]

$$DSC = \frac{2Tp}{2Tp+Fp+Fn} \tag{III.2}$$

**IV.3. Partition coefficient (Vpc)**

In 1981, Bezdek introduced the partition coefficient, an index of class validity depending on the degrees of belonging Uik. The partition coefficient Vpc is defined as:

$$\mathrm{Vpc}(\mathrm{U},\mathrm{C}) = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{C}} \mathrm{U}_{\mathrm{ik}}^{2} \tag{III.3}$$

In the case where the value of Vpc is equal to one, indicating that all classes are well delimited. In the other case where Vpc equals zero, each element belongs to all classes with the same degree of membership and Vpc indicating that the classification is as blurry as possible. Therefore, with the increase in the quality of the classification, the value of Vpc also increases.

**IV.4. Similarity accuracy :**

Similarity accuracy is an evaluation metric commonly used in image segmentation tasks to measure the performance and quality of segmentation algorithms or models. It provides a quantitative assessment of how well the predicted segmentation aligns with the ground truth or reference segmentation.

Similarity accuracy evaluates the overlap or agreement between the predicted segmentation and the ground truth segmentation by considering the similarity between the two sets of pixels or regions. It takes into account both the number of correctly classified pixels or regions (true positives) and the number of incorrectly classified pixels or regions (false positives and false negatives).

higher values of similarity accuracy indicate better segmentation performance, with values closer to 100 indicating more accurate and precise segmentations. These metrics provide a useful and intuitive way to evaluate and compare different segmentation methods, enabling researchers and practitioners to assess the quality of their algorithms and make informed decisions for further improvements.

The Similarity accuracy (SA) is defined as:

$$\text{simillarity accuraccy} = \frac{\text{well classified pixel counts}}{\text{total number of pixels}} \times 100 \qquad \textbf{III.4}$$

## V. Segmentation results of the synthetic images :

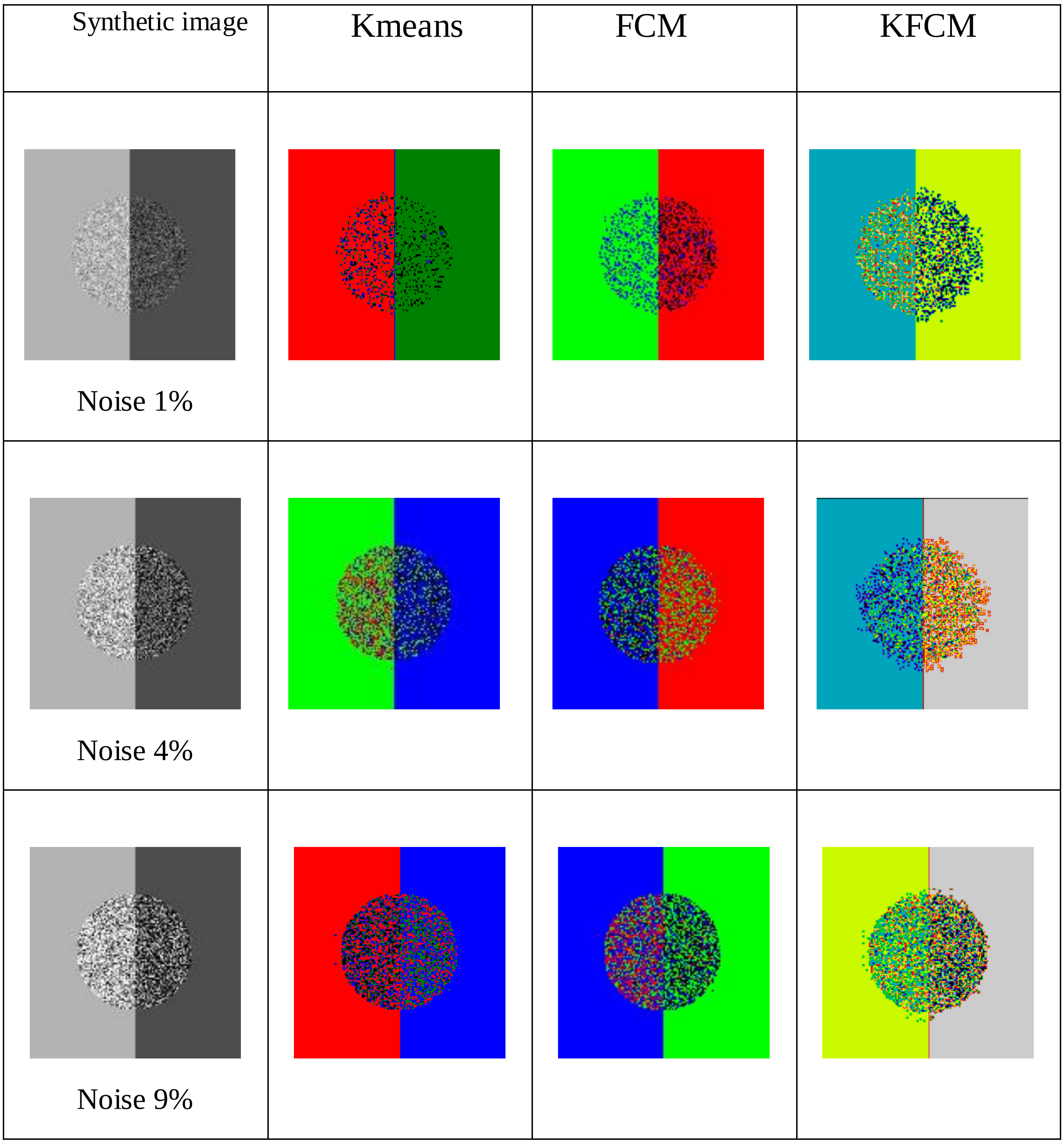



Figure III. 4: segmentation results of synthetic images

### V.1 Discusion:

It is noted that the quality of the result obtained is greatly degraded for a high noise level (9%). Like the K-means algorithm, FCM is very sensitive to noise.

In the case of this image, a similarity evaluation is necessary. Several criteria are used in the literature for the evaluation of classification methods. It is obvious that each criterion is adapted to the method used and the type of data that is manipulated. We have chosen to use the jaccard index and dice coefficient and partition coefficient and similarity accuracy.

## V.2. Synthetic images results:

### V.2.1. Synthetic Images with 1% Noise:

Table III. 1:Results of the segmentation on a 1% noise synthetic image

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.99 | 0.99 | 0.99 |
| Dice coefficient (%) | 0.99 | 0.99 | 0.99 |
| Partition coefficient (%) | 0.97 | 0.99 | 0.99 |
| Simillarity accuracy (%) | 99.49 | 99.96 | 99.98 |

### V.2.2. Synthetic Images with 4% Noise:

Table III. 2:Results of the segmentation on a 4% noise synthetic image

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.98 | 0.99 | 0.99 |
| Dice coefficient (%) | 0.99 | 0.99 | 0.99 |
| Partition coefficient (%) | 0.97 | 0.98 | 0.99 |
| Simillarity accuracy (%) | 98.90 | 99.37 | 99.38 |

### V.2.3 Synthetic Images with 9% Noise:

Table III. 3:Results of the segmentation on a 9% noise synthetic image

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.97 | 0.98 | 0.98 |
| Dice coefficient (%) | 0.99 | 0.99 | 0.99 |
| Partition coefficient (%) | 0.96 | 0.97 | 0.98 |
| Simillarity accuracy (%) | 98.32 | 98.79 | 98.79 |

**V.2.4. Discusion:**

this table compares the performance of three clustering algorithms (K-means, FCM, and KFCM) on synthetic images with different levels of noise. The evaluation metrics used include the Jaccard index, Dice coefficient, Partition coefficient, and SA (Sensitivity Accuracy)

For the jaccard index we can see that For each algorithm, as the noise level increases from 1% to 9%, the Jaccard index decreases consistently. This indicates that the clustering results become less similar to the ground truth as the noise increases. KFCM consistently outperforms FCM and K-means, achieving higher Jaccard index values for all noise levels. This suggests that KFCM is more effective in handling noisy data and producing clusters that align better with the ground truth.

Similar to the Jaccard index, the Dice coefficient also decreases as the noise level increases for all algorithms. KFCM again demonstrates superior performance, generaly achieving higher Dice coefficient values compared to FCM and K-means.

The partition coefficient measures the similarity between the generated clusters and the ground truth clusters. Similar to the previous metrics, as the noise level increases, the partition coefficient decreases for all algorithms.KFCM consistently achieves higher partition coefficient values compared to FCM and K-means, indicating better cluster quality.

Similarity accuracy measures the accuracy of cluster assignments.As expected, as the noise level increases, the SA decreases for all algorithms.KFCM maintains higher SA values compared to FCM and K-means, implying more accurate cluster assignments.

## VI. Segmentation of the MRI images :

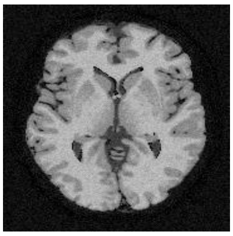
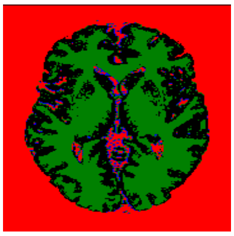
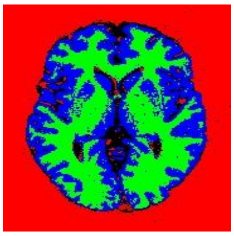
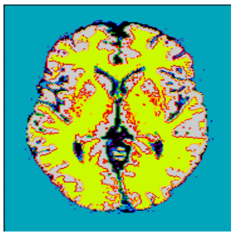
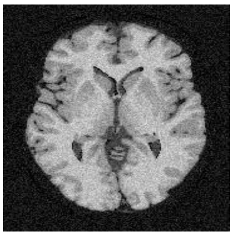
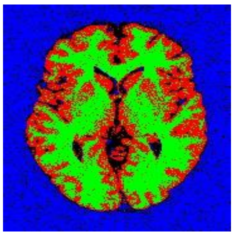
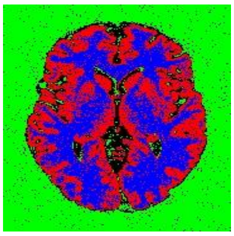
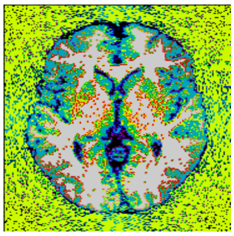
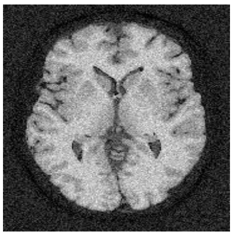
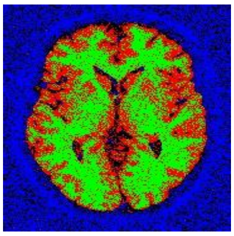
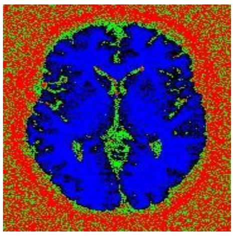
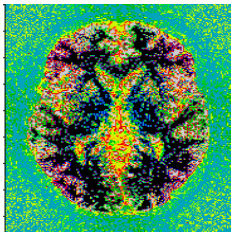
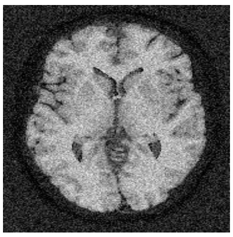
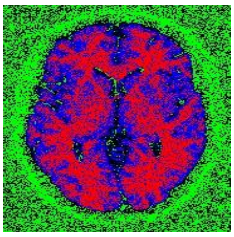
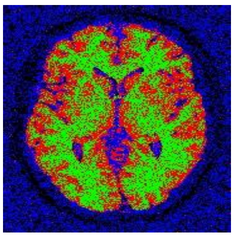
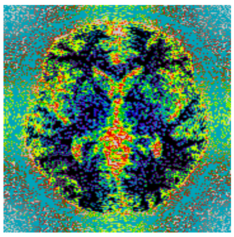
| MRI images | Kmeans | FCM | KFCM |
|---|---|---|---|
|  Noise 3% |  |  |  |
|  Noise 5% |  |  |  |
|  Noise 7% |  |  |  |
|  Noise 9% |  |  |  |

Figure III. 5:Results of the Segmentation performed on MRI images

### VI.1. Discussion :

It is clear that as the noise level increases, the classification quality deteriorates. But it is not easy to decide which is the best result. So to better compare the results, we calculate the Jaccard index and dice coefficient and partition coefficient and similarity accuracy for the different partitions obtained and then we can decide which is the best algorithm and the best segmentation result.

## VII. MRI images results :

Table III. 4:MRI images segmentation Results for 3% noise

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.89 | 0.93 | 0.97 |
| Dice coefficient (%) | 0.94 | 0.96 | 0.97 |
| Partition coefficient (%) | 0.89 | 0.93 | 0.97 |
| Simillarity accuracy (%) | 95.65 | 97.30 | 97.33 |

Table III. 5:MRI images segmentation Results for 5% noise

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.87 | 0.90 | 0.96 |
| Dice coefficient (%) | 0.92 | 0.95 | 0.96 |
| Partition coefficient (%) | 0.82 | 0.90 | 0.95 |
| Simillarity accuracy (%) | 95.10 | 95.56 | 96.58 |

Table III. 6:MRI images segmentation Results for 7% noise

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.83 | 0.86 | 0.95 |
| Dice coefficient (%) | 0.91 | 0.92 | 0.95 |
| Partition coefficient (%) | 0.81 | 0.82 | 0.95 |
| Simillarity accuracy (%) | 94.91 | 96.90 | 95.44 |

Table III. 7:MRI images segmentation Results for 9% noise

| | Kmeans | FCM | KFCM |
|---|---|---|---|
| Jaccard index (%) | 0.80 | 0.82 | 0.94 |
| Dice coefficient (%) | 0.89 | 0.90 | 0.93 |
| Partition coefficient (%) | 0.80 | 0.82 | 0.94 |
| Simillarity accuracy (%) | 93.16 | 94.34 | 94.39 |

### VII.1. Comparison :

Table III. 8:Comparison Results for the three algorithms

| Algorithm | K-means | | | FCM | | | KFCM | | |
|---|---|---|---|---|---|---|---|---|---|
| Noise | 3% | 5% | 7% | 3% | 5% | 7% | 3% | 5% | 7% |
| Jaccard Index (%) | 0.89 | 0.87 | 0.83 | 0.93 | 0.90 | 0.86 | 0.97 | 0.96 | 0.95 |
| Dice Coeff (%) | 0.94 | 0.92 | 0.91 | 0.96 | 0.95 | 0.92 | 0.97 | 0.96 | 0.95 |
| Partition coeff (%) | 0.89 | 0.82 | 0.81 | 0.93 | 0.90 | 0.82 | 0.97 | 0.95 | 0.95 |
| SA (%) | 95.65 | 95.10 | 94.91 | 97.30 | 95.56 | 96.90 | 97.33 | 96.58 | 95.44 |

### VII.2. Discussion :

From the provided table, we can observe the performance of three clustering algorithms: K-means, Fuzzy C-means (FCM), and Kernelized Fuzzy C-means (KFCM), on three sets of MRI images. The evaluation metrics used to measure the quality of clustering are the Jaccard index, Dice coefficient, Partition coefficient, and the Sensitivity and Accuracy (SA).

we can see all three algorithms experience a decrease in their performance across all evaluation metrics. This is expected since noise can introduce uncertainty and disrupt the clustering process.

Comparing the algorithms, we observe that KFCM performs relatively well in the presence of noise, achieving a Jaccard index of 0.97, 0.96 and 0.95 for noise levels of 3%, 5%, and 7% respectively. FCM and K-means exhibit slightly lower performance, with FCM achieving Jaccard indices of 0.93, 0.90, and 0.86, and K-means achieving Jaccard indices of 0.89, 0.87 and 0.83 for the respective noise levels.

Moving on to the Dice coefficient and Partition coefficient, we can make similar observations. KFCM outperforms FCM and K-means in both metrics, with higher values indicating better clustering results.

The only surprised result is that the Similarity accuracy as we can see the algorithm FCM has a small advantage than KFCM . This may be due to the way these two algorithms model the similarity between pixels or objects in the image. FCM is based on the minimization of Euclidean distances, while KFCM uses a kernel function to measure similarity.

based on the provided results, we can say that KFCM demonstrates better robustness to noise compared to FCM and K-means, achieving higher values across most of evaluation metrics. However, it is important to note that these results may vary depending on the specific dataset and the choice of parameters for each algorithm.

# VIII. Classes separation for segmented MRI image :

## VIII.1. Separated classes results of MRI image  with 3% noise results:

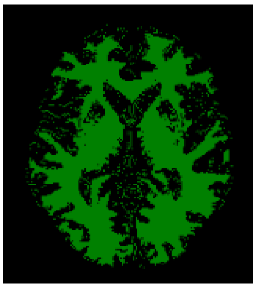
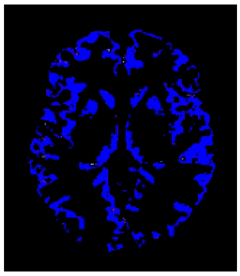
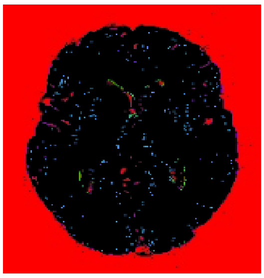
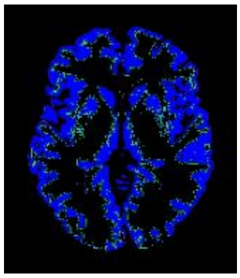
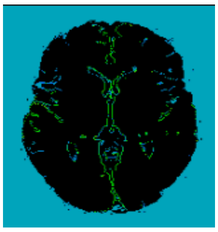
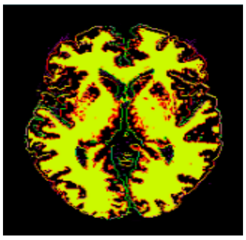
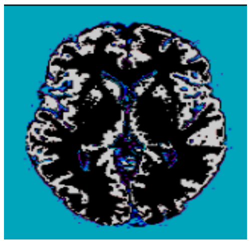
| | 3%noise | | |
| --- | --- | --- | --- |
| | CSF | WM | GM |
| K means |  |  |  |
| FCM |  |  |  |
| KFCM |  |  |  |

Figure III. 6 :Separated classes of MRI image  with 3% noise results

**VIII.2. Separated classes results of MRI image with 7% noise results:**

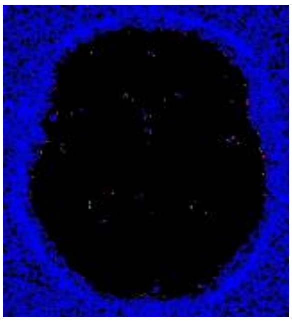
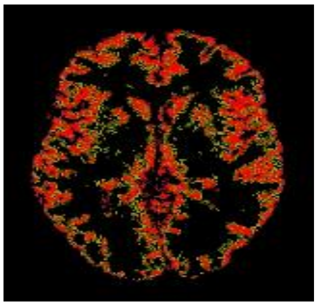
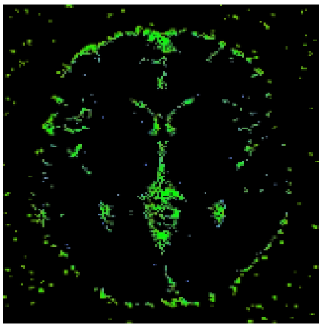
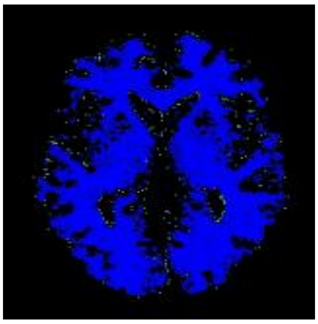
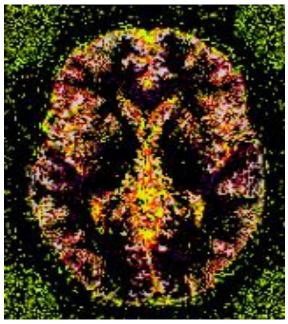
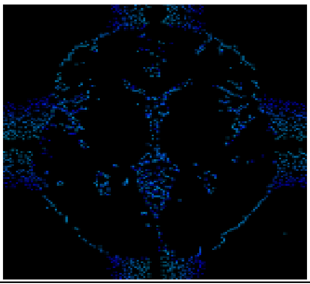
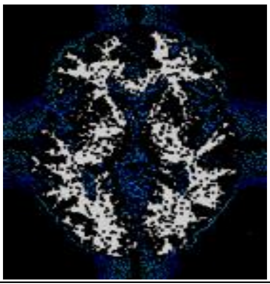
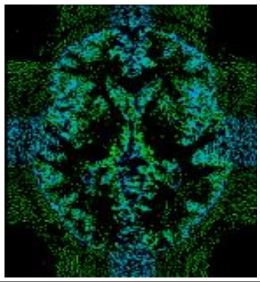
| 7%noise | | | |
|---|---|---|---|
| | CSF | WM | GM |
| K means |  |  |  |
| FCM |  |  |  |
| KFCM |  |  |  |

Figure III. 7:Separated classes of MRI image with 7% noise results

### VIII.3. Separated classes results of MRI image with 9% noise results:

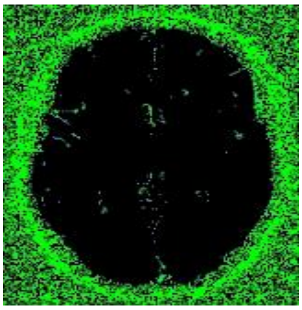
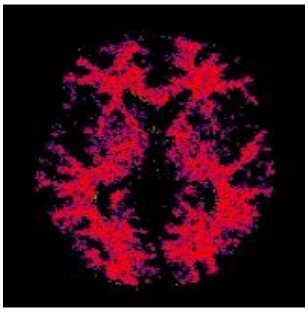
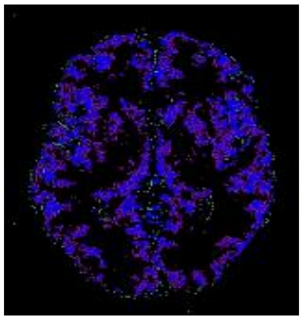
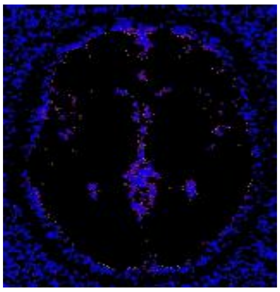
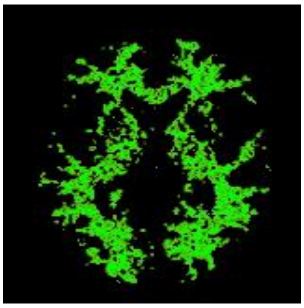
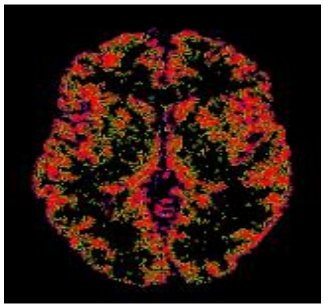
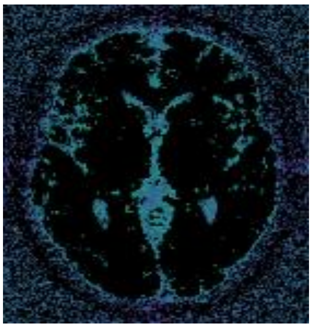
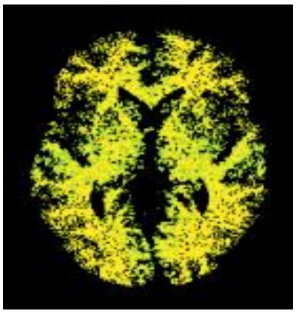
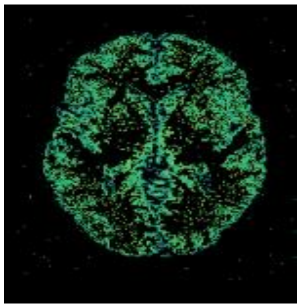
| 9%noise | | | |
|---|---|---|---|
| | CSF | WM | GM |
| K means |  |  |  |
| FCM |  |  |  |
| KFCM |  |  |  |

Figure III. 8:Separated classes of MRI image with 9% noise results

### VIII.4. Discussion:

When applying K-means, FCM, KFCM to MRI images with different levels of noise, the quality of segmentation results can vary. As the noise level increases (from 3 to 9), the accuracy of segmentation generally decreases. Higher noise levels introduce additional variability in intensity values, leading to misclassification and overlapping clusters. Consequently, the boundaries between CSF, white matter, and Gray matter become less distinct.

To confirm this and better compare the performance of the three algorithms we need to make an evaluation to the segmentation of separated classes.

## IX. Evaluation of the segmentation for the separated classes:

### IX.1. Separated classes of MRI image with 3% Noise:

Table III. 9 :Separated classes segmentation Results for 3% noise

| Algorithm | K-means | | | FCM | | | KFCM | | |
|---|---|---|---|---|---|---|---|---|---|
| Brain Classes | CSF | WM | GM | CSF | WM | GM | CSF | WM | GM |
| Jaccard Index (%) | 0.34 | 0.54 | 0.60 | 0.40 | 0.46 | 0.46 | 0.60 | 0.63 | 0.79 |
| Dice Coeff (%) | 0.36 | 0.57 | 0.63 | 0.43 | 0.47 | 0.49 | 0.62 | 0.66 | 0.81 |
| Partition coeff (%) | 0.33 | 0.48 | 0.58 | 0.40 | 0.42 | 0.42 | 0.60 | 0.61 | 0.75 |
| SA (%) | 35 | 54 | 60 | 43 | 46 | 47 | 63 | 60 | 78 |

### IX.2. Separated classes of MRI image with 7% noise :

Table III. 10:Separated classes segmentation Results for 7% noise

| Algorithm | K-means | | | FCM | | | KFCM | | |
|---|---|---|---|---|---|---|---|---|---|
| Brain Classes | CSF | WM | GM | CSF | WM | GM | CSF | WM | GM |
| Jaccard Index (%) | 0.27 | 0.44 | 0.48 | 0.33 | 0.53 | 0.58 | 0.52 | 0.57 | 0.59 |
| Dice Coeff (%) | 0.31 | 0.47 | 0.49 | 0.33 | 0.53 | 0.59 | 0.53 | 0.59 | 0.63 |
| Partition coeff (%) | 0.29 | 0.40 | 0.43 | 0.28 | 0.49 | 0.55 | 0.54 | 0.44 | 0.57 |
| SA (%) | 36 | 44 | 48 | 33 | 53 | 57 | 57 | 49 | 59 |

### IX.3. Separated classes of MRI image with 9% noise :

Table III. 11:Separated classes segmentation Results for 9% noise

| Algorithm | K-means | | | FCM | | | KFCM | | |
|---|---|---|---|---|---|---|---|---|---|
| Brain Classes | CSF | WM | GM | CSF | WM | GM | CSF | WM | GM |
| Jaccard Index (%) | 0.26 | 0.49 | 0.47 | 0.31 | 0.41 | 0.52 | 0.44 | 0.66 | 0.46 |
| Dice Coeff (%) | 0.31 | 0.52 | 0.47 | 0.31 | 0.44 | 0.56 | 0.49 | 0.63 | 0.48 |
| Partition coeff (%) | 0.29 | 0.44 | 0.40 | 0.38 | 0.36 | 0.46 | 0.44 | 0.63 | 0.41 |
| SA (%) | 34 | 50 | 47 | 37 | 42 | 52 | 52 | 65 | 47 |

### IX.4. Discussion :

From the previous tables we notice that KFCM (Kernel-based Fuzzy C-Means) outperforms both K-means and FCM (Fuzzy C-Means) algorithms in terms of overall performance. However, as the level of noise increases, the performance of all three algorithms decreases.

When comparing the classes, we notice a decrease in segmentation accuracy for the CSF (Cerebrospinal Fluid) class across all three algorithms as the noise level increases. This implies that the noise interferes with the separation process of this particular class, leading to less accurate segmentation.

In contrast, the WM (White Matter) class shows a decrease in segmentation accuracy as well, but K-means algorithm demonstrates better performance in maintaining the accuracy of WM segmentation when the noise level is increased.

On the other hand, the GM (Gray Matter) class appears to be the most stable among the three classes, as its segmentation accuracy shows minimal decrease across all three algorithms, regardless of the noise level. This suggests that the GM class is less affected by noise interference in the segmentation process.

According to those results , we can say that KFCM is better and more robust to noise compared to FCM and K-means, achieving higher values across most of evaluation metrics.

## X. Comparison between FCM and KFCM:

**Non-linear Clustering**:

KFCM incorporates a kernel function that allows non-linear transformations of the input data. This enables KFCM to capture complex and non-linear relationships among data points, leading to more accurate clustering in scenarios where linear separation is not sufficient. FCM, on the other hand, assumes linear relationships between data points and clusters.

**Flexibility in Cluster Shape**:

By introducing kernel transformations, KFCM can handle clusters of various shapes and sizes. It is not limited to the assumption of spherical clusters, as in the case of K-means or FCM. This flexibility allows KFCM to capture clusters with irregular shapes, elongated structures, or overlapping regions.

**Robustness to Noise and Outliers**:

FCM is generally more robust to noise and outliers compared to K-means due to fuzzy membership values. However, KFCM further improves robustness by incorporating kernel transformations. The kernel function can help suppress the influence of noisy data or outliers, allowing for more reliable and accurate clustering results.

Overall, the results suggest that KFCM performs better than FCM and K-means in terms of all the evaluated metrics. It exhibits higher similarity to the ground truth clusters and produces more accurate and reliable cluster assignments, especially in the presence of higher levels of noise. This indicates that KFCM is a more robust algorithm for clustering tasks, particularly when dealing with noisy data.

## XI . Conclusion :

In summary, K-means is a straightforward algorithm that assumes crisp partitions, while FCM extends this by allowing fuzzy membership values. KFCM further enhances FCM by incorporating kernel transformations to handle non-linear data relationships. FCM and KFCM are generally more flexible, robust to noise, and can handle clusters of varying shapes and sizes. However, they also come with increased computational complexity compared to K-means. The choice between the algorithms depends on the specific characteristics of the data and the desired clustering goals.

# General conclusion

# General conclusion

In conclusion, work has focused on the utilization of the Kernel Fuzzy C-Means (KFCM) method for the segmentation of medical images, with the primary objective of aiding doctors in their clinical practice.

The research findings have significant potential in assisting medical professionals by enabling accurate segmentation of medical images, facilitating precise identification and analysis of anatomical structures, lesions, and other crucial regions of interest. This, in turn, can contribute to improved diagnosis, treatment planning, and patient monitoring.

The superiority of the Kernel Fuzzy C-Means algorithm over other segmentation methods, such as K-Means and Fuzzy C- Means has been demonstrated throughout this work, While the mentioned algorithms showed less efficiency in segmentation accuracy as the noise level increased.

The adoption of the KFCM algorithm displays its potential as a robust and accurate tool for segmenting medical images.

One of the notable strengths of the KFCM algorithm lies in its high noise resistance, enabling robust segmentation of medical images even in the presence of noise or interference, making it a powerful tool for medical image analysis. By leveraging kernel methods, the algorithm enables non-linear transformations, leading to more precise and flexible segmentation results.

Furthermore, the incorporation of fuzzy logic within the KFCM algorithm effectively addresses uncertainties and variations in image intensities, enhancing the overall segmentation process.

Future works can focus on exploring the integration of advanced techniques, such as deep leaning, The KFCM algorithm can take advantage of the powerful capabilities of neural networks to improve the accuracy and precision of its segmentation results. Deep learning enables the algorithm to learn complex patterns and features directly from the data, leading to more reliable and detailed segmentation.

# Bibliography

[1] Mr.manseur abdelghani, Cours Traitement d'Images ,master 1 electronique des systems embarqué ,Université Kasdi Merbah Ouargla, 2022

[2] Rafael C. Gonzalez .Digital Image. Processing. Third Edition. University of Tennessee. Richard E. Woods. NledData Interactive. 2007

[3] Mr. Rohit Verma, Dr. Jahid Ali, A Comparative Study of Various Types of Image Noise and Efficient Noise Removal Techniques , international Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 10, October 2013

[4] Caponetti, L.; Castellano, G.; Corsini, V. MR Brain Image Segmentation: A Framework to Compare Different Clustering Techniques. Information 2017, *8*, 138. https://doi.org/10.3390/info8040138 (Accessed on May 2023 )

[5] Mohammdian-khoshnoud, M., Soltanian, A.R., Dehghan, A. et al. Optimization of fuzzy c-means (FCM) clustering in cytology image segmentation using the gray wolf algorithm. BMC Mol and Cell Biol **23**, 9 (2022). https://doi.org/10.1186/s12860-022-00408-7 (Accessed on May 2023 )

[6] Dao-Qiang Zhang, Song-Can Chen,A novel kernelized fuzzy C-means algorithm with application in medical image segmentation,Artificial Intelligence in Medicine,Volume 32, Issue 1,2004,

[7] Marcelo R.P. Ferreira a,b , Francisco de A.T. de Carvalho b , Kernel fuzzy c-means with automatic variable weighting, 16 May 2013

[8] Kalam, Rehna & Thomas, Ciza & M, Abdul. Gaussian Kernel Based Fuzzy C-Means Clustering Algorithm for Image Segmentation. Computer Science & Information Technology. 6. 47-56. 10.5121/csit.2016.60405. (2016).

[9] Rezki Aulia Putri , Kernel Based Fuzzy C-Means Clustering for Chronic Sinusitis Classification , et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 546 052060

[10] Further Generalizations of the Jaccard Index, Luciano da Fontoura Costa , S˜ao Carlos Institute of Physics – DFCM/USP 10th Oct 2021

[11] Ortega, Joaquín & Almanza-Ortega, Nelva & Vega-Villalobos, Andrea & Pazos-Rangel, Rodolfo & Zavala-Diaz, José Crispin & Martínez-Rebollar, Alicia. (2019). The K-Means Algorithm Evolution. 10.5772/intechopen.85447.

[12] Pérez-Ortega J, Nely Almanza-Ortega N, Vega-Villalobos A, et al. (2020) The K-Means Algorithm Evolution. Introduction to Data Science and Machine Learning. IntechOpen. DOI: 10.5772/intechopen.85447.

[13] D'Hondt Frédéric, El Khayati Brahim, « Etude de méthodes de Clustering pour segmentation d'images en couleurs », Faculté Polytechnique de Mons, 5ème Electricité, Certificat Applicatifs Multimédia, Janvier 2005

[14] D'Hondt Frédéric, El Khayati Brahim, « Etude de méthodes de Clustering pour segmentation d'images en couleurs », Faculté Polytechnique de Mons, 5ème Electricité, Certificat Applicatifs Multimédia , Janvier 2005

[15] Dr Vivek Jaglan .Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm Vikas Thada Research Scholar Department of Computer Science and Engineering Dr. K.N.M University,Newai, Rajasthan, 2013

[16] Tyagi, Vipin. (2018). Understanding Digital Image Processing. 10.1201/9781315123905.

[17] Pixabay.com ( Public Domain Images ) ( Accessed on february 2023 )

[18] http://www.bic.mni.mcgill.ca/brainweb/ ( Accessed on May 2023)

[19] J.-P.COCQUEREZ et S.PHILIPP, « Analyse d'images : filtrage et segmentation », page3 -266, Paris, (février 2003)

[20] S. K. M Shadekul Islam .(2023). Introduction of Medical Imaging Modalities. Research and Development Department, Pioneer Alpha, Dhaka, Bangladesh

[21] Anatoly Zolotkov, Système de Coordonnées Cartésiennes, https://www.translatorscafe.com/unit-converter/fr-FR/calculator/two-points-distance/ (Accessed on the March 2023)

[22] Zouaoui Hakima, Moussaoui Abdelouahab, « Clustering par fusion floue de données appliqué à la segmentation d'images IRM cérébrales ».

## Abstract:

This study explores how the Kernel Fuzzy C-Means (KFCM) method can be used for medical image segmentation. The effectiveness of the KFCM method evaluated by comparing it with other well-known methods like fuzzy c-means and k-means. To ensure accurate assessment, proven evaluation metrics such as the Jaccard index and the Dice coefficient used for objective analysis.

The experimental results demonstrate that the KFCM method is successful in accurately identifying different parts of the brain in various medical images. By combining the Fuzzy C-Means algorithm with kernel functions, the KFCM method improves the accuracy of clustering, resulting in more precise and dependable segmentation outcomes.

This research contributes to the progress of medical image segmentation techniques, highlighting the promising capabilities of the KFCM method. Overall, the findings suggest that the KFCM method has the potential to be a valuable tool in medical image analysis, providing better results and advancing the field of medical diagnostics.


Keywords: KFCM ,K-means , FCM, Medical image segmentation, Clustering, Jaccard, Dice

## Résumé :

Cette étude explore comment la méthode Kernel Fuzzy C-Means (KFCM) peut être utilisée pour la segmentation d'images médicales. L'efficacité de la méthode KFCM est évaluée en la comparant à d'autres méthodes bien connues telles que fuzzy c-means et k-means.

Pour garantir une évaluation précise, des métriques d'évaluation qui ont fait leurs preuves telles que l'indice de Jaccard et le coefficient de Dice sont utilisées pour une analyse objective.

Les résultats expérimentaux démontrent que la méthode KFCM parvient à identifier avec précision différentes parties du cerveau dans diverses images médicales. En combinant l'algorithme Fuzzy C-Means avec des fonctions Kernels, la méthode KFCM améliore la précision du regroupement, ce qui se traduit par des résultats de segmentation plus précis et fiables.

Cette recherche contribue à l'avancement des techniques de segmentation d'images médicales, mettant en évidence les capacités prometteuses de la méthode KFCM. Globalement, les résultats suggèrent que la méthode KFCM a le potentiel d'être un outil précieux dans l'analyse d'images médicales, offrant de meilleurs résultats et faisant progresser le domaine du diagnostic médical.


Mots clés : KFCM ,FCM , K-means,segmentation d'images médicales, l'indice de Jaccard

**ملخص:**

تناولت هذه الدراسة استخدام طريقة Kernel Fuzzy C-Means (KFCM) لتجزئة الصور الطبية.

يتم تقييم فعالية طريقة KFCM عن طريق مقارنتها مع طرق معروفة أخرى مثل طريقة fuzzy c-means وطريقة k-means لضمان التقييم الدقيق، يتم استخدام مقاييس التقييم المثبتة مثل مؤشر جاكارد ومعامل دايس للتحليل الموضوعي.

تشير النتائج التجريبية إلى أن طريقة KFCM ناجحة في تحديد أجزاء مختلفة من الدماغ في الصور الطبية المختلفة بدقة من خلال دمج خوارزمية Fuzzy C-Means مع وظائف النواة، تحسن طريقة KFCMدقة التجميع، مما يؤدي إلى نتائج تجزئة أكثر دقة وموثوقية.

يساهم هذا البحث في تقديم تقنيات تجزئة الصور الطبية، مسلطة الضوء على قدرات طريقة KFCM الواعدة. بشكل عام، تشير النتائج إلى أن طريقة KFCM لديها إمكانات لتكون أداة قيمة في تحليل الصور الطبية، مما يوفر نتائج أفضل ويعزز مجال التشخيص الطبي.


الكلمات المفتاحية : FCM,K-means KFCM, مؤشر جاكارد, معامل دايس,صور الطبية