الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة مصطفى اسطمبولي معسكر
Université Mustapha Stambouli de Mascara

Faculté des Sciences Exactes
Département Informatique

Thèse de Doctorat Es-science

Spécialité : Informatique

Intitulé

Segmentation d'Images médicales basée sur les Approches de l'Intelligence Artificielle

Présentée par : Mokhtari chakir

Devant le jury :

Président	Salem Mohammed	Professeur	Université de Mascara
Directeur de thèse	Meftah Boudjelal	Professeur	Université de Mascara
Examinateur	Keskes Nabil	Professeur	ESI de Sidi Bel Abbès
Examinateur	Benyahya Kadda	MCA	Université de Saida
Examinateur	Zennaki Mahmoud	MCA	UST'Oran
Examinateur	Rebbah Mohammed	Professeur	Université de Mascara

Année universitaire 2025-2026

الجمهورية الجزائرية الديمقراطية الشعبية
The Democratic and Popular Algerian Republic
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة مصطفى اسطمبولي معسكر
Mustapha Stambouli University of Mascara

Faculty of Exact Sciences
Department of Computer Science

Doctoral Thesis in Science

Specialty: Computer Science

Title

Medical Image Segmentation Based on Artificial Intelligence Approaches

Presented by: Mokhtari chakir

Jury members:

President	Salem Mohammed	Professor	Mascara University
Thesis Advisor	Meftah Boudjelal	Professor	Mascara University
Examiner	Keskes Nabil	Professor	ESI of Sidi Bel Abbès
Examiner	Benyahya Kadda	MCA	Saida University
Examiner	Zennaki Mahmoud	MCA	UST'Oran
Examiner	Rebbah Mohammed	Professor	Mascara University

Academic year: 2025-2026

Dedications

This thesis is dedicated to the memory of my beloved parents, whose love, sacrifices, and unwavering belief in my potential have been a constant source of inspiration. They may no longer be here with me, but their spirit lives on in every page of this work.

To my dear wife, for her endless patience, encouragement, and understanding. Her support has been the anchor that kept me grounded throughout this journey. Also to the memory of her beloved parents and their sisters and brother.

To my wonderful children, Mohammed Firas, Fatima Zohra, Mustapha Nadir, and Amina Alaa, who are my greatest motivation. May this work inspire you to pursue your dreams with passion and perseverance.

To my sisters and brothers, for their encouragement and support.

To my second family Yahiani, Hadj Abderezak, God bless him.

Finally, to my loyal friends, who stood by me and celebrated every small victory along the way.

Acknowledgment

It is a pleasure to express my deepest gratitude to all the individuals who contributed to the successful completion of this thesis.

My sincere appreciation goes to my supervisor, Pr. Meftah Boudjelal, for his invaluable guidance, unwavering support, and constructive feedback throughout this research. His expertise and encouragement were instrumental in shaping this work.

I would also like to extend my thanks to the members of the examining committee for graciously accepting to evaluate this thesis:

I would first like to express my deep gratitude to Pr. Salem Mohammed, full professor at Mustapha Stambouli University in Mascara, for having done me the honor of chairing the jury for my thesis defense.

I extend my deepest gratitude to Mr. Keskes Nabil, Full Professor at Higher School of Computer Science of Sidi Bel Abbès; Benyahia Kadda, Associate professor at Doctor Moulay Taher University of Saida; and Mr. Zennaki Mahmoud, Associate professor at Mohammed Boudiaf University of Sciences and Technology in Oran, for graciously accepting to serve on the jury and dedicating their time to review and enhance my thesis.

I wish to also express my sincere appreciation to Mr. Rebbah Mohammed, Full Professors at Mustapha Stambouli University in Mascara, for his valuable interest in this work and for graciously accepting to serve on the jury.

Your insightful comments and suggestions were highly beneficial and helped improve the quality of this work.

Finally, a special acknowledgment is due to Dr. Debakla Mohammed, whose support and assistance were of great help during my research journey.

Summary

Medical image segmentation is a critical preprocessing step in computer-aided diagnosis, treatment planning, and biomedical research. While Fuzzy C-Means (FCM) clustering is a widely adopted technique for this task due to its ability to handle inherent ambiguities in medical data, its performance is highly sensitive to initial parameters and is prone to convergence to local optima.

This thesis presents a comprehensive approach to overcoming these limitations through two primary contributions. First, we provide an overview of medical imaging and the fundamental challenges of segmentation. We then detail traditional clustering-based methods, with a specific focus on the FCM algorithm, outlining its strengths and well-documented limitations. To address these limitations, we explore bio-inspired optimization metaheuristics as a powerful strategy for guiding the clustering process.

The core contribution of this work is the novel hybridization of the Artificial Bee Colony (ABC) algorithm with FCM. The proposed method focuses on the simultaneous optimization of the crucial FCM parameters: primarily the number of cluster centers and their values and the optimization of the objective function by escaping to the local optima, to achieve a superior and more robust segmentation outcome. The effectiveness of this hybrid ABC-FCM approach is rigorously validated through experiments on both simulated brain MRI and real clinical MRI brain images. Results demonstrate a significant improvement in segmentation accuracy and convergence behavior compared to standard FCM and other optimization-enhanced variants.

The second major contribution is the development of a new cluster validity index (CVI) to automatically determine the optimal number of segments. This index is designed to enhance the separation metric of the IMI index by incorporating a measure based on Kullback-Leibler (KL) divergence, which better captures the statistical distance between fuzzy clusters. Experimental results confirm that the proposed KL-based CVI outperforms existing indices in accurately identifying the true number of clusters in both synthetic and complex medical imagery.

This thesis offers significant advancements in AI-driven medical image segmentation by introducing an optimized clustering framework and a more robust validation metric, both contributing to higher diagnostic reliability.

Keywords:

Medical Image Segmentation, Fuzzy C-Means (FCM), Artificial Bee Colony (ABC) Algorithm, Metaheuristic Optimization, Magnetic Resonance Imaging (MRI), Cluster Validity Index (CVI), Kullback-Leibler Divergence.

ملخص

تعد تقسيمة الصور الطبية خطوة معالجة مسبقة حاسمة في التشخيص بمساعدة الكمبيوتر، وتخطيط العلاج، والبحث الطبي الحيوي. على الرغم من أن خوارزمية التجميع الضبابي (Fuzzy C-Means - FCM) هي تقنية معتمدة على نطاق واسع لهذه المهمة بسبب قدرتها على التعامل مع الغموض الكامن في البيانات الطبية، إلا أن أداءها حساس للغاية للمعاملات الأولية وعرضة للتقارب نحو القيم المثلى المحلية.

تقدم هذه الأطروحة نهجًا شاملاً للتغلب على هذه القيود من خلال مساهمتين رئيسيتين. أولاً، نقدم نظرة عامة على التصوير الطبي والتحديات الأساسية للتقسيمة. ثم نقوم بتفصيل الطرق التقليدية القائمة على التجميع، مع التركيز بشكل خاص على خوارزمية FCM، مع التأكيد على نقاط قوتها وقيودها الموثقة جيدًا. لمعالجة هذه القيود، نستكشف metaheuristicsالتحسين المستوحاة من الطبيعة (bio-inspired) كإستراتيجية قوية لتوجيه عملية التجميع.

المساهمة الأساسية لهذا العمل هي التهجين بين خوارزمية مستعمرة النحل الاصطناعية Artificial Bee Colony) (ABC -وخوارزمية FCM. تركز الطريقة المقترحة على التحسين المتزامن للمعا ملات الحاسمة لـ FCM مثل عدد مراكز التجمعات وقيمها وتحسين دالة الهدف من خلال الهروب من القيم المثلى المحلية، لتحقيق نتيجة تقسيمة متفوقة أكثر. يتم التحقق من فعالية نهج ABC-FCM الهجين هذا بدقة عبر تجارب على صور الرنين المغناطيسي للدماغ بالمحاكاة والسريرية الحقيقية. تظهر النتائج تحسنًا كبيرًا في دقة التقسيم.

تتمثل المساهمة الرئيسية الثانية في تطوير مؤشر جديد لصلاحية التجمعات (CVI) يهدف إلى تحديد العدد الأمثل للتجمعات بشكل تلقائي. وقد تم تصميم هذا المؤشر لتحسين مقياس الفصل الخاص بمؤشرIMI ، من خلال دمج مقياس يعتمد على تباعد Kullback-Leibler، الذي يوفّر تمثيلًا أدق للمسافة الإحصائية بين التجمعات الضبابية. وتؤكد النتائج التجريبية أن مؤشر المقترح، المبني على تباعد Kullback-Leibler، يتفوق على المؤشرات الحالية من حيث الدقة في تحديد العدد الحقيقي للتجمعات، سواء في الصور التركيبية أو الصور الطبية المعقدة.

تُقدّم هذه الأطروحة إسهامات بارزة في مجال تقسيم الصور الطبية المدعوم بالذكاء الاصطناعي، من خلال اقتراح إطار تجميع محسّن، ومقياس تحقق أكثر قوة، مما يساهم في تعزيز موثوقية التشخيص بشكل ملحوظ.

الكلمات المفتاحية:

تقسيم الصور الطبية، خوارزمية Fuzzy C-Means (FCM) ، خوارزمية مستعمرة النحل الاصطناعية (ABC) ، التحسين باستخدام الميتا-استراتيجيات، التصوير بالرنين المغناطيسي (MRI/IRM) ، مؤشر صلاحية التجمعات (CVI) ، تباعد Kullback-.Leibler

Résumé

La segmentation d'images médicales constitue une étape prétraitement cruciale pour le diagnostic assisté par ordinateur, la planification des traitements et la recherche biomédicale. Bien que la méthode de clustering Fuzzy C-Means (FCM) soit largement utilisée pour cette tâche grâce à sa capacité à gérer les ambiguïtés inhérentes aux données médicales, sa performance est très sensible aux paramètres initiaux et tend à converger vers des optimums locaux.
Cette thèse présente une approche complète pour surmonter ces limitations à travers deux contributions principales. Premièrement, nous fournissons une vue d'ensemble de l'imagerie médicale et des défis fondamentaux de la segmentation. Nous détaillons ensuite les méthodes traditionnelles basées sur le clustering, en nous concentrant particulièrement sur l'algorithme FCM, en soulignant ses forces et ses limitations bien documentées. Pour répondre à ces limitations, nous explorons les métaheuristiques d'optimisation bio-inspirées comme stratégie puissante pour guider le processus de clustering.
La contribution principale de ce travail est l'hybridation novatrice de l'algorithme Artificial Bee Colony (ABC) avec le FCM. La méthode proposée se concentre sur l'optimisation simultanée des paramètres cruciaux du FCM : principalement le nombre de centres de clusters et leurs valeurs, et l'optimisation de la fonction objective en échappant aux optimums locaux, pour obtenir un résultat de segmentation supérieur et plus robuste. L'efficacité de cette approche hybride ABC-FCM est rigoureusement validée par des expérimentations sur des images cérébrales IRM simulées et réelles. Les résultats démontrent une amélioration significative de la précision de segmentation et du comportement de convergence comparé au FCM standard et à d'autres variants optimisés.
La deuxième contribution majeure est le développement d'un nouvel indice de validité de clusters (CVI) pour déterminer automatiquement le nombre optimal de segments. Cet indice est conçu pour améliorer la métrique de séparation de l'indice IMI en incorporant une mesure basée sur la divergence de Kullback-Leibler (KL), qui capture mieux la distance statistique entre les clusters flous. Les résultats expérimentaux confirment que le CVI proposé basé sur KL surpasse les indices existants en identifiant avec précision le nombre réel de clusters dans des images tant synthétiques que médicales complexes.
Cette thèse offre des avancées significatives dans la segmentation d'images médicales pilotée par l'IA en introduisant un framework de clustering optimisé et une métrique de validation plus robuste, contribuant tous deux à une fiabilité diagnostique accrue.

Mots-clés :
Segmentation d'Images Médicales, Fuzzy C-Means (FCM), Algorithme Artificial Bee Colony (ABC), Optimisation par Métaheuristiques, Imagerie par Résonance Magnétique (IRM), Indice de Validité de Clusters (CVI), Divergence de Kullback-Leibler.

Contents

Chapter 2: Medical Image Segmentation

Chapter 3: FCM Optimisation based on Bio-Inspired Methods

Chapter 4: Hybrid FCM-ABC Method for Medical Image Segmentation

Chapter 5: Fuzzy Validity Index based on Kullback-Leibler Divergence

List of Figures

List of Tables

General Introduction

General Introduction

Medical image segmentation serves as a cornerstone of modern healthcare, enabling precise delineation of anatomical structures and pathological regions across diverse imaging modalities, from Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) to Positron Emission Tomography (PET) and ultrasound. With the exponential growth of these medical imaging modalities, the demand for accurate, efficient, and automated segmentation techniques has intensified. However, medical images present inherent challenges, including (1) noise and artifacts due to sensor limitation, patient emotion and modality-specific distortions like MRI bias fields degrade image quality, (2) intensity inhomogeneity: Tissue intensity overlaps (gray/white matter in MRIs) complicate boundary detection, (3) partial volume effects: Voxels containing mixed tissues due to limited resolution blur structural boundaries, and (4) anatomical variability: Inter-patient diversity and pathological anomalies (tumors, lesions) demand adaptive solutions, such factors significantly impact the performance of automated segmentation algorithms and highlight the need for advanced AI solutions capable of handling these complexities. The limitations of conventional segmentation methods like thresholding, region growing, and edge detection become particularly apparent when dealing with such imperfect data. These traditional approaches often rely on rigid assumptions that fail to account for the uncertainty and variability inherent in medical images, motivating the development of more sophisticated techniques. However, traditional methods remain relevant for specific applications and as preprocessing steps in modern segmentation pipelines.

The advent of Artificial Intelligence (AI) has transformed the field by introducing data-driven paradigms that learn intricate patterns directly from imaging data. Among these, Machine Learning Approach (MLA) and Deep Learning Approach (DLA).

Machine learning, including supervised and unsupervised techniques, has significantly advanced medical image segmentation by offering more robust, data-driven approaches compared to traditional methods. Supervised methods like support vector machines (SVMs) and random forests use labeled datasets to learn complex patterns, improving segmentation accuracy. However, their success depends on high-quality annotated data, which is costly and time-consuming to produce, and they often struggle to generalize across different imaging protocols or populations. Unsupervised methods, such as k-means clustering, Gaussian mixture models (GMMs), and fuzzy c-means (FCM), group pixels based on similarity without labeled data, making them useful for exploratory analysis. However, they lack precision for clinical applications due to reliance on low-level features and sensitivity to noise and artifacts. Both approaches face challenges like intensity inhomogeneities, noise, class imbalance, and high computational costs, which can degrade performance and limit scalability. While machine learning remains relevant in specific applications and hybrid pipelines, its challenges highlight the need for continued innovation in medical image segmentation.

The rise of deep learning, particularly convolutional neural networks (CNNs), revolutionized brain MRI segmentation by enabling automatic learning of hierarchical features from raw data. Architectures like U-Net, with its contracting and expansive paths connected by skip connections, excelled in capturing fine details and achieving state-of-the-art results. Fully convolutional networks (FCNs) further advanced the field by enabling end-to-end, pixel-wise segmentation without handcrafted features. However, deep learning methods face challenges, including the need for large, high-quality annotated datasets, which are costly and time-consuming to produce. Limited dataset diversity can hinder model performance and generalization, even with data augmentation. Additionally, the high computational cost of training, especially for volumetric data, poses scalability and accessibility issues, particularly in resource-constrained settings. Despite these limitations, deep learning remains a transformative approach in medical image segmentation.

In this thesis, we advocate for the hybridization of the Fuzzy C-Means (FCM) method applied to medical image segmentation, positioning it as a compelling alternative to purely machine learning (MLA) and deep learning (DLA) approaches. While MLA and DLA methods have revolutionized medical image segmentation with their ability to learn complex patterns and achieve state-of-the-art results, they come with significant challenges, including the need for large annotated datasets, high computational costs, and limited interpretability. In contrast, we investigate these advanced AI-driven methodologies for medical image segmentation which bridge the gap between classical machine learning and bio-inspired optimization methods. We focus on enhancing Fuzzy C-Means (FCM) clustering, a prominent soft segmentation technique for handling pixel-level uncertainty, via bio-inspired optimization methods. While FCM is particularly well-suited for medical imaging due to its ability to handle the inherent ambiguity and uncertainty in tissue boundaries and unlike traditional hard clustering methods, it allows pixels to belong to multiple clusters with varying degrees of membership, reflecting the partial volume effect often observed in medical image. This flexibility makes FCM highly effective for segmenting object of interest. However, traditional FCM presents serious limitations which can degrade its performance in complex medical image modalities datasets.

Firstly, it needs the right number of clusters which is not available in most cases.

Secondly, it is very sensitive to initialization, deferent cluster centers initialization can lead to deferent clustering results.

Thirdly, due to the principle of the iterative optimization of a cost function, it is strongly sensitive to the problems of local minima. These challenges can lead to suboptimal segmentation results, particularly in complex MRI datasets with intensity inhomogeneities or overlapping tissue distributions.

Shortcomings that we aim to address.

So to address these shortcomings, we propose a novel hybrid framework integrating the Artificial Bee Colony (ABC) algorithm which is a swarm intelligence metaheuristic inspired by honeybee foraging behavior, to automate the optimization of *all* FCM parameters like cluster centroids initialization and their optimal values, their number and membership matrix *simultaneously*.

Key innovations include:

- ABC-driven FCM optimization: ABC's global search capability mitigates FCM's convergence to suboptimal local minima, enhancing robustness in complex datasets like brain MRIs with lesions or tumors.
- KL divergence-based cluster validity index: A new evaluation metric leveraging Kullback-Leibler (KL) divergence to quantify segmentation quality by measuring the statistical divergence between pixel intensity distributions and cluster prototypes. This addresses the bias of traditional indices (like Xie-Beni, Partition Coefficient) toward specific cluster geometries.

Expected Contributions

1. Automated parameter tuning: Elimination of manual intervention via ABC's adaptive optimization, improving reproducibility.
2. Objective evaluation: The KL Divergence-based validity index provides a statistically grounded measure for comparing segmentation outcomes across various modalities.

Validation

Experiments on public neuroimaging datasets (Simulated and real brain data sets) will demonstrate the framework's superiority over conventional FCM and other Hybrid methods in metrics like Jaccard similarity, Partition Coefficient and Entropy and Davies-Bouldin index

Thesis Structure

The thesis is organized into five chapters, each addressing a critical aspect of AI-driven medical image segmentation:

Chapter 1: Overview of Medical Imaging

This chapter provides a comprehensive review of medical imaging modalities, their underlying physics, clinical applications, and the technical challenges they create for automated analysis. We examine the characteristics of major imaging techniques including MRI, CT, PET, ultrasound, and X-ray, with particular emphasis on their diagnostic and therapeutic roles in clinical practice. Each modality has unique advantages - MRI offers superior soft tissue contrast without ionizing radiation, CT provides rapid acquisition of high-resolution anatomical structures, and PET delivers functional metabolic information. However, they also present modality-specific artifacts and limitations that must be addressed. MRI suffers from intensity

inhomogeneity and susceptibility artifacts; CT has limited soft tissue contrast and uses ionizing radiation, while PET exhibits poor spatial resolution and high noise levels.

Chapter 2: Medical Image Segmentation

We review in this chapter fundamental and state-of-the-art segmentation methods, including thresholding, region-based, edge-based, and machine learning approaches. Special attention is given to clustering-based techniques, highlighting their advantages and limitations in medical imaging.

Chapter 3: FCM Optimization based on Bio-Inspired methods

This chapter delves into fuzzy clustering theory, focusing on the FCM algorithm and its variants (spatial FCM, kernel FCM, ..). We then explore bio-inspired optimization methods such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC) and their applications in enhancing clustering performance.

Chapter 4: Hybrid FCM-ABC Method for Medical Image Segmentation

Here, we present our primary contribution: an ABC-optimized FCM framework where cluster centroids, membership degrees, and number of clusters are simultaneously tuned for optimal brain MRI segmentation. Experimental results demonstrate superior performance compared to conventional FCM and other hybrid approaches.

Chapter 5: Fuzzy Validity Index Based on Kullback-Leibler Divergence

We propose an innovative cluster validity measure leveraging Kullback-Leibler divergence to quantify segmentation quality more effectively. The proposed index is rigorously evaluated against existing metrics, demonstrating improved robustness in assessing fuzzy partitions.

Research Contributions

- Optimized FCM via ABC: A fully automated FCM optimization framework that eliminates manual parameter tuning and enhances segmentation accuracy.

- New Validity Index: A mathematically sound fuzzy validity measure based on KL divergence for objective evaluation of segmentation results.

- Clinical Applicability: Validation on real brain MRI datasets, showcasing the method’s potential in neuro-imaging applications such as tumor detection and tissue analysis.

Conclusion and Future Work

The thesis concludes by summarizing key findings, discussing clinical implications, and outlining future research directions, including the integration of deep learning with fuzzy clustering and extensions to multi-modal medical image segmentation.

By advancing AI-driven segmentation techniques, this work contributes to more reliable and automated medical image analysis, ultimately supporting improved diagnostic precision and personalized treatment strategies.

Chapter 1 : Overview of Medical Imaging

1. Introduction

Medical imaging is a cornerstone of modern healthcare, playing an essential role in diagnosis, treatment planning, and monitoring of diseases and conditions. It encompasses a wide range of techniques that allow healthcare professionals to visualize the internal structures of the body, aiding in the detection and understanding of various diseases. These images serve as a window into the human body, providing insight that is crucial for accurate diagnosis, surgical planning, and ongoing patient care.

One of the core principles behind medical imaging is its non-invasive nature, which allows clinicians to examine patients' internal organs and structures without the need for surgery. This non-invasive approach not only minimizes patient discomfort but also reduces the risk of complications, making it a preferred choice for diagnostic purposes.

The medical imaging begins with the discovery of X-rays in 1895 by German physicist Wilhelm Conrad Roentgen. This groundbreaking discovery revolutionized the medical field, enabling doctors to see inside the human body for the first time. Roentgen's work led to the creation of the first X-ray images, which were initially used to examine broken bones. Over the decades, the technology progressed, expanding into new fields like mammography and fluoroscopy.

The next major leap in medical imaging came with the advent of Computed Tomography (CT) in the early 1970s. Godfrey Hounsfield and Alan Cormack were awarded the Nobel Prize in Physiology or Medicine in 1979 for their work in developing this imaging technology. CT combined traditional X-ray technology with computers, enabling the creation of cross-sectional images or slices of the body, providing far more detailed information than a standard X-ray.

In the 1980s, Magnetic Resonance Imaging (MRI) emerged as a promising imaging technique. MRI technology uses strong magnetic fields and radio waves to generate detailed images of soft tissues, making it particularly useful for brain, spinal cord, and joint imaging. Unlike X-ray and CT, MRI does not rely on radiation, making it a safer option for certain patient populations.

In the years that followed, Nuclear Medicine introduced new possibilities for functional imaging. Techniques such as Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) revolutionized the way physicians could visualize and assess how organs and tissues are functioning, not just their structural appearance [Anthony et al., 2013]

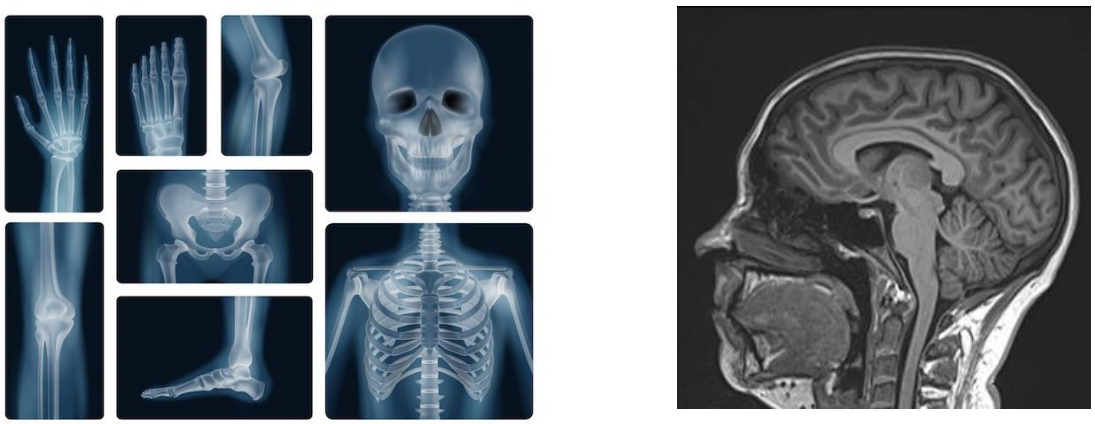
This chapter provides an overview of various medical imaging modalities, such as X-ray, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, and Nuclear Medicine, each with its own set of advantages, limitations, and specific clinical applications.

2. Types of Medical Imaging

Medical imaging techniques can be broadly classified into several categories based on the technology used, the type of information they provide, and their specific clinical applications. Below is an overview of the most widely used modalities

2.1. X-ray and Computed Tomography (CT)

X-ray and CT imaging are the most commonly used medical imaging techniques. They utilize radiation to produce images of the body's internal structures (cf.fig.1.1).



(a) (b)

Figure 1.1: X-ray (a) and CT (b) images of different parts of the human body

X-ray imaging is a widely utilized diagnostic tool, particularly effective for visualizing bone fractures, joint dislocations, and dental examinations. It is also commonly employed in screening for lung infections, such as pneumonia, and in mammography for breast cancer detection [Carlton et al., 2013]. On the other hand, CT scans utilize X-ray technology combined with advanced computer processing to create highly detailed images of the body's internal structures. By stacking multiple cross-sectional X-ray slices, CT scans can generate 3D images, offering superior visualization of organs, bones, and blood vessels. This makes CT scans particularly valuable in complex diagnostic scenarios where detailed imaging is critical [Seeram, 2015].

2.1.1. Clinical Applications of X-ray

X-ray imaging plays a critical role in various clinical applications due to its ability to provide quick and detailed images of internal structures. It is particularly effective for diagnosing fractures, dislocations, and joint abnormalities, making it a cornerstone in orthopedics and trauma care. In chest imaging, X-rays are widely used to evaluate conditions such as pneumonia, tuberculosis, lung cancer, and heart failure, offering valuable insights into lung and heart health. Dental X-rays are another essential application, enabling dentists to detect cavities, abscesses, impacted teeth, and bone loss, which are crucial for maintaining oral health. Additionally, mammography, a specialized form of X-ray, is a key tool in breast cancer screening, helping to identify tumors, calcifications, and other abnormalities that may

indicate early stages of cancer. These diverse applications underscore the versatility and importance of X-ray imaging in modern medicine.

2.1.2. Clinical Applications of CT

CT has a broad range of applications in both diagnosis and treatment planning. Some of the major uses include:

- Trauma and Emergency Medicine: CT imaging is vital in emergency medicine, particularly for trauma, enabling rapid assessment of head, neck, spine, chest, and abdominal injuries. It diagnoses intracranial hemorrhages, brain swelling, skull fractures, internal bleeding, organ ruptures, and spinal fractures with precision.
- Cancer Staging: this technique of medical imaging is a cornerstone in oncology, playing a vital role in the detection, staging, and monitoring of cancer. It is widely used to detect tumors in various organs, including the lungs, liver, pancreas, and colon, providing detailed information about their size, location, and characteristics. CT is also instrumental in cancer staging, helping to determine the extent of metastasis and assess lymph node involvement, which is critical for treatment planning. Additionally, CT scans are frequently employed to monitor the effectiveness of cancer treatments, such as chemotherapy or radiation therapy, by tracking changes in tumor size over time. Its ability to deliver precise, high-resolution images makes CT an indispensable tool in the fight against cancer.
- Cardiac Imaging: in cardiology, it provides detailed visualization of the heart and vascular system, aiding in the diagnosis and management of various conditions. One of its primary applications is in evaluating coronary artery disease, where coronary CT angiography (CTA) offers a non-invasive method to image the coronary arteries, detect atherosclerotic plaques, and identify narrowing or blockages. Additionally, CT is instrumental in assessing cardiac abnormalities such as aortic aneurysms, pulmonary embolism, and congenital heart defects.
- Neurological Imaging: CT imaging is a critical tool for evaluating conditions affecting the brain and spinal cord, providing rapid and detailed insights for diagnosis and treatment. In cases of stroke, CT scans are essential for distinguishing between ischemic strokes, hemorrhages, and brain edema, enabling timely and appropriate interventions. For brain tumors, CT helps visualize abnormalities such as tumors, cysts, and abscesses, assisting clinicians in planning effective treatment strategies. Additionally, CT is valuable in diagnosing hydrocephalus by detecting abnormal fluid accumulation in the brain's ventricles.
- Pulmonary Imaging: it is widely utilized for detailed evaluation of the lungs and airways, playing a key role in diagnosing critical conditions. CT pulmonary angiography (CTPA) is the gold standard for detecting pulmonary

embolism (PE), a life-threatening blockage in the lung's arteries. It is also essential for assessing chronic lung diseases, such as chronic obstructive pulmonary disease (COPD) and interstitial lung disease, providing insights into disease progression and management. Additionally, CT is crucial for the early detection and staging of lung cancer, enabling timely intervention and treatment planning.

- Musculoskeletal Imaging: CT imaging is indispensable for evaluating complex bone fractures and joint-related conditions, offering detailed and precise visualization that aids in accurate diagnosis and treatment planning. It is particularly valuable for assessing intricate fractures in areas such as the spine, pelvis, and long bones, where traditional imaging methods may fall short. Additionally, CT is highly effective in diagnosing joint abnormalities, including arthritis, bone infections, and inflammatory conditions like rheumatoid arthritis.

2.1.3. Advanced CT techniques

Several advanced CT techniques have been developed to enhance image quality, minimize radiation exposure, and improve diagnostic accuracy.

1- Dual-Energy CT utilizes two distinct X-ray energy levels to capture images, enabling better tissue differentiation. This technique is particularly useful for detecting tumors, urinary stones, and vascular conditions, while also reducing the need for contrast material.

2- Iterative Reconstruction (IR) is a mathematical approach that lowers radiation doses without compromising image quality. By reducing image noise, IR enhances diagnostic precision, making it especially beneficial for pediatric imaging and high-risk patients.

3- Cardiac CT is a specialized application that provides detailed visualization of the heart and coronary vessels. It is widely used to evaluate coronary artery disease, identify coronary artery anomalies, and assist in pre-operative planning for heart surgery. These advancements have significantly expanded the capabilities of CT imaging, making it safer and more effective for a wide range of clinical applications.

2.1.4. Advantages and limitations of X-rays and CT imaging techniques

Both X-ray imaging and CT scans come with distinct advantages and limitations. X-rays are known for their speed, cost-effectiveness, and widespread availability, making them well diagnostic tool in many medical settings. However, they expose patients to ionizing radiation, which carries potential health risks, and are less effective at imaging soft tissues compared to modalities like MRI or ultrasound. CT scans, while providing more detailed and comprehensive images, also involve higher doses of ionizing radiation. Although the radiation dose is generally considered safe for most patients, repeated exposure over time can increase the risk of radiation-induced conditions.

Therefore, the use of both imaging techniques requires careful consideration, ensuring that the diagnostic benefits outweigh the potential risks, especially for patients who may need frequent imaging:

- Radiation exposure: While CT is highly effective, it involves a significantly higher dose of radiation compared to conventional X-rays, which raises concerns about the cumulative effects of repeated scans.

- Cost: CT scans are more expensive compared to traditional X-ray imaging.

- Limited soft tissue contrast: Although CT is better at imaging soft tissues than traditional X-ray, MRI is often preferred for certain soft tissue evaluation (brain, spinal cord, and muscles).

2.2 Magnetic Resonance Imaging (MRI)

MRI utilizes strong magnetic fields and radiofrequency waves to produce images. The human body is largely made up of water, and water molecules consist of hydrogen atoms, which are highly responsive to magnetic fields. When placed in a magnetic field, hydrogen atoms align in a certain direction. Radiofrequency pulses are then used to temporarily disturb this alignment. As the hydrogen atoms return to their original state, they emit signals, which are detected and used to create an image [Viallon et al., 2015].

MRI generates highly detailed images of the body's internal structures, especially soft tissues such as the brain, spinal cord, muscles, and organs. It is particularly valuable for neurological, orthopedic, and cardiovascular imaging [Arnold et al., 2023] (cf.fig.1.2).

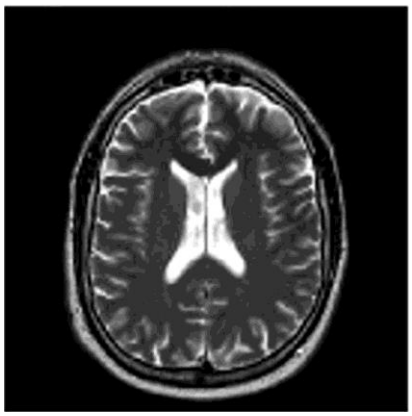


Figure 1.2: example of Brain MRI image

2.2.1. Clinical Applications of MRI

MRI is versatile and plays a crucial role in diagnosing a wide variety of conditions due to its ability to generate high-resolution images of soft tissues. Some key applications of MRI include:

- Neurological Imaging: MRI is widely regarded as the gold standard for imaging the brain and spinal cord due to its superior soft tissue contrast and detailed visualization capabilities. For the brain, MRI is highly effective in detecting a range of conditions, including brain tumors, stroke, multiple sclerosis, epilepsy, and neurodegenerative diseases such as Alzheimer's disease. Its ability to provide high-resolution images of brain structures makes it indispensable for accurate diagnosis and monitoring. Similarly, MRI excels in evaluating spinal cord conditions, such as spinal cord injuries, herniated discs, and spinal stenosis. It offers precise imaging that is critical for treatment planning and surgical interventions. Overall, MRI's advanced imaging capabilities make it a vital tool in neurology and neurosurgery.
- Musculoskeletal Imaging: MRI is highly effective for soft tissue imaging, making it an invaluable tool for assessing muscles, ligaments, tendons, and cartilage. It is widely used to diagnose sports-related injuries, such as ligament tears, muscle strains, and joint abnormalities, offering detailed insights that guide treatment and rehabilitation. Additionally, MRI plays a key role in evaluating osteoarthritis by visualizing cartilage wear and changes in bone structure. This helps clinicians understand the progression of the disease and develop appropriate management strategies. With its exceptional ability to provide high-resolution images of soft tissues and joints, MRI is a cornerstone in orthopedics and musculoskeletal imaging.
- Cardiac Imaging: It is commonly used to assess cardiac function, diagnose myocardial infarction (heart attack), and evaluate heart valve diseases, offering precise information that aids in treatment planning. Additionally, Magnetic Resonance Angiography (MRA) is a specialized MRI technique used to visualize blood vessels non-invasively, without the need for ionizing radiation or invasive procedures. MRA is particularly useful for detecting vascular abnormalities, such as aneurysms, blockages, or malformations, making it a valuable tool in cardiovascular imaging. Together, these applications highlight MRI's versatility and importance in diagnosing and managing heart and vascular conditions.
- Oncology: MRI is a critical imaging modality in oncology, particularly for detecting and staging soft tissue cancers such as those in the liver, prostate, and breast. Its superior soft tissue contrast allows for the identification of tumors that may not be easily visible on other imaging techniques, making it an invaluable tool for early cancer detection. It plays a key role in tumor staging by providing detailed information about the size, location, and extent of tumor involvement in surrounding tissues. This information is essential for developing effective treatment plans and monitoring disease progression.
- Abdominal and Pelvic Imaging: MRI is a highly effective imaging modality for evaluating abdominal and pelvic organs, providing exceptional detail for diagnosing a wide range of conditions. In the abdomen, MRI is particularly useful for imaging the liver, kidneys, pancreas, and spleen, making it invaluable for detecting conditions such as cirrhosis, tumors, and inflammatory diseases. In the pelvis, MRI is widely used to assess the uterus, ovaries, and prostate, offering detailed visualization that aids in the

detection of cancers, structural abnormalities, and other pathologies [Jin 2021]. Its ability to produce high-resolution, multi-planar images without ionizing radiation makes MRI a preferred choice for diagnosing and managing diseases affecting these critical organs.

2.2.2. Advanced MRI Techniques

Over the years, several advanced MRI techniques have been developed to improve image quality, reduce scan times, and enhance diagnostic capabilities.

- Functional MRI (fMRI): measures brain activity by detecting changes in blood flow. It is commonly used in neuroscience research and pre-surgical mapping of brain function. fMRI can help identify regions of the brain responsible for tasks like movement, speech, and sensory perception.

- Diffusion-Weighted Imaging (DWI): is an MRI technique that measures the movement of water molecules in tissues. It is particularly useful for detecting acute ischemic stroke, as ischemic tissue shows restricted water movement.

- Magnetic Resonance Spectroscopy (MRS): provides information about the chemical composition of tissues, allowing for the detection of metabolic changes in diseases like cancer or neurodegenerative disorders.

2.2.3. Advantages and limitations of MRI

MRI offers significant advantages as a non-invasive diagnostic tool that does not rely on ionizing radiation, making it a safer option for patients who require frequent imaging. It excels in producing high-resolution images, particularly of soft tissues, which makes it invaluable for diagnosing conditions affecting the brain, spinal cord, muscles, and joints. This level of detail is often unmatched by other imaging modalities, allowing for precise detection and evaluation of abnormalities.

However, MRI also has notable limitations. The procedure is generally more expensive and time-consuming compared to other imaging techniques, which can limit its accessibility and practicality in urgent situations. Additionally, MRI is not suitable for patients with certain implants, such as pacemakers or metal devices, due to the strong magnetic fields involved. These constraints highlight the importance of carefully considering patient-specific factors when choosing MRI as a diagnostic tool.

2.3. Ultrasound Imaging

Ultrasound imaging, also known as sonography, is a non-invasive medical imaging technique that uses high-frequency sound waves to create real-time images of the inside of the body. Unlike other imaging methods, ultrasound does not involve the use of ionizing radiation, making it a safe and widely used modality for diagnostic purposes, especially in obstetrics, cardiology, and musculoskeletal imaging.

Ultrasound imaging works based on the principle of sound wave reflection [Sehmbi & Perlas, 2022]. High-frequency sound waves (usually above 20 kHz) are emitted from a transducer and travel through the body. These sound waves encounter tissues of different densities, and part of the wave is reflected back to the transducer. The time it takes for the sound waves to return and the intensity of the reflected waves are used to generate an image (cf.fig.1.3).

Figure 1.3: Ultrasound image

2.3.1. Clinical Applications of Ultrasound

Ultrasound is used across a wide variety of medical specialties, with particular advantages in imaging soft tissues and organs, monitoring pregnancies, and guiding certain medical procedures.

- Obstetrics and Gynecology: Ultrasound is a vital tool in women's health, widely used in obstetrics to monitor fetal development, assess growth, heartbeat, and placental health. It also detects abnormalities like ectopic pregnancies, multiples, and birth defects. Beyond pregnancy, ultrasound aids in fertility and gynecological care by evaluating the ovaries, uterus, and fallopian tubes, helping diagnose conditions like fibroids and endometriosis. Its non-invasive nature and real-time imaging make it indispensable in modern medicine.
- Cardiovascular Imaging: Echocardiography is a specialized ultrasound technique that evaluates the heart's structure and function, providing detailed images of its chambers, valves, and blood flow. It is essential for diagnosing conditions like heart failure, valvular heart disease, cardiomyopathies, and congenital heart defects, offering critical insights into cardiac health through a non-invasive approach. Complementing this, Doppler ultrasound assesses blood flow within the body's vessels by measuring its speed and direction, helping identify abnormalities such as blockages, stenosis, aneurysms, and venous thrombosis. Particularly valuable for diagnosing peripheral artery disease and deep vein thrombosis, Doppler ultrasound is a cornerstone in managing circulatory disorders. Together, echocardiography and Doppler ultrasound

significantly enhance the ability to diagnose and treat cardiovascular and vascular conditions effectively, making them indispensable tools in modern medicine.

- Musculoskeletal Imaging: this technique is a highly effective tool for evaluating joints and soft tissues, providing detailed images of tendons, ligaments, muscles, and bursae. It is commonly used to diagnose conditions such as tendonitis, ligament tears, muscle strains, and arthritis. By offering real-time, high-resolution imaging, ultrasound allows healthcare providers to accurately assess the extent of injuries or inflammation, facilitating targeted treatment plans. In addition to diagnostics, ultrasound plays a crucial role in guiding minimally invasive procedures. It is frequently used to assist with joint injections, ensuring precise delivery of medications such as corticosteroids for conditions like arthritis. Ultrasound guidance is also invaluable for biopsy procedures and the aspiration of fluid from cysts or joints, whether for diagnostic testing or therapeutic relief. This combination of diagnostic accuracy and procedural precision makes ultrasound an essential tool in musculoskeletal and interventional medicine.
- Abdominal and Pelvic Imaging: it is widely used for evaluating abdominal organs, including the liver, gallbladder, kidneys, spleen, and pancreas. It is particularly effective in detecting conditions such as gallstones, liver disease, renal abnormalities, and tumors, providing critical diagnostic information without the need for invasive procedures. Its ability to deliver real-time, high-resolution images makes it a first-line tool for assessing abdominal health. In pelvic imaging, ultrasound plays an equally important role for both men and women. It is used to visualize the bladder, prostate, and reproductive organs, aiding in the diagnosis of conditions like benign prostatic hyperplasia (BPH) in men and ovarian cysts or fibroids in women. This non-invasive technique offers valuable insights into pelvic health, enabling accurate diagnosis and effective management of a wide range of conditions. Ultrasound's adaptability and safety make it an essential tool in both abdominal and pelvic diagnostics.
- Thyroid Imaging: Ultrasound is commonly used to evaluate the thyroid gland for nodules, cysts, or cancer. It helps determine the size, texture, and blood flow characteristics of thyroid abnormalities, often guiding further biopsy or treatment decisions.
- Guiding Procedures: it is also an invaluable tool for guiding minimally invasive procedures, offering real-time visualization to ensure accuracy and safety. It is frequently used to direct needle biopsies, fluid aspirations, and injections, particularly in deep or hard-to-reach areas. This precision reduces the risk of complications and improves diagnostic and therapeutic outcomes. Additionally, ultrasound plays a critical role in guiding the drainage of abscesses or cysts, allowing healthcare providers to safely remove fluid while avoiding damage to surrounding tissues. Its ability to provide clear, real-time imaging makes ultrasound an essential asset in interventional medicine, enhancing the effectiveness of a wide range of procedures.

2.3.2 Advanced Ultrasound Techniques

Several advanced ultrasound techniques have been developed to enhance diagnostic capabilities:

- Doppler Ultrasound: is used to measure blood flow within blood vessels. It is often used to detect vascular conditions such as deep vein thrombosis (DVT), venous insufficiency, vascular malformations, and atherosclerosis. It can also evaluate the flow of blood through the heart, detecting conditions such as valvular insufficiency or stenosis.
- 3D and 4D Ultrasound: 3D Ultrasound technique captures a series of 2D images and reconstructs them into 3D volumes. It is commonly used in obstetrics to generate more detailed images of the fetus. While 4D Ultrasound allows for real-time visualization of fetal movements in the womb.
- Elastography: is an advanced ultrasound technique that measures the stiffness of tissues. It is commonly used to assess liver stiffness as an indicator of fibrosis or cirrhosis and can be applied to other tissues to detect abnormalities such as tumors.

2.3.3. Advantages and limitations of Ultrasound

Ultrasound imaging offers several advantages as a diagnostic tool, including being non-invasive, safe, and free from ionizing radiation, making it a preferred option for various patient populations, including pregnant women. Its ability to provide real-time imaging allows for dynamic assessment of moving structures, such as the heart or blood flow, which is particularly useful in procedures like echocardiograms or guiding biopsies. This immediacy and safety profile make ultrasound a versatile and widely used imaging modality.

However, ultrasound has certain limitations. While it excels in visualizing soft tissues, it struggles to image bones and air-filled organs, such as the lungs. The quality of ultrasound images can vary significantly depending on the operator's skill and experience, making it operator-dependent. It is also less effective for imaging deeper tissues, particularly in obese patients or those with a larger body habitus. Furthermore, the presence of gas or air, such as in the intestines, can interfere with sound wave transmission, limiting its effectiveness in certain areas. Despite these limitations, ultrasound remains a versatile and invaluable tool in modern medicine.

2.4. Nuclear Medicine

Nuclear Medicine is a branch of medical imaging that uses radioactive substances (radiopharmaceuticals) to diagnose and treat diseases. Unlike traditional imaging techniques that visualize structures, nuclear medicine primarily provides functional and metabolic information about organs and tissues. By detecting the radiation emitted from radiopharmaceuticals, nuclear medicine helps assess organ function, detect disease, and guide therapy [Iskandrian & Hage, 2024].

The most common types of nuclear medicine imaging include Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET). Both

techniques offer insight into the biological processes within the body, providing critical information for disease diagnosis and treatment planning.

2.4.1. Single Photon Emission Computed Tomography (SPECT)

SPECT is a nuclear medicine imaging technique that uses gamma-emitting radiopharmaceuticals to create 3D images of the body (cf.fig.1.4). It is a versatile diagnostic tool widely utilized across multiple medical fields. In cardiac imaging, SPECT is commonly employed to evaluate coronary artery disease, myocardial infarction, and overall heart function. It provides detailed visualization of areas with reduced blood flow, helping to identify ischemic tissue and guide treatment decisions. In neurology, it is used to diagnose conditions such as epilepsy, Alzheimer's disease, and Parkinson's disease by assessing brain activity and blood flow patterns [Verger et al., 2021]. In oncology, SPECT plays a crucial role in detecting tumors and evaluating the spread of cancer by highlighting areas of abnormal tissue metabolism and growth.

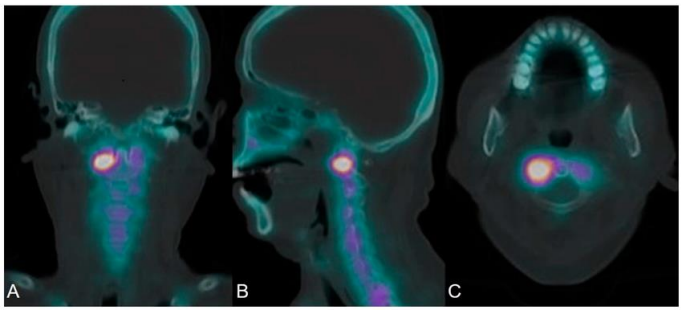


Figure: 1.4: example of SPECT image.

2.4.2. Positron Emission Tomagraphy (PET)

PET is a sophisticated imaging technique that uses positron-emitting radiopharmaceuticals to produce high-resolution, 3D images of metabolic activity in the body (cf.fig.1.5). PET scans are particularly effective in detecting cancer, evaluating brain function, and assessing cardiac and neurological diseases.

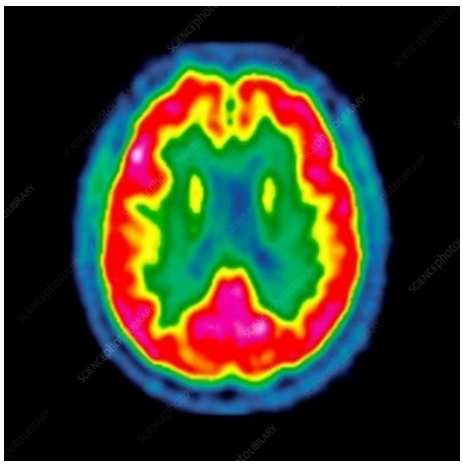


Figure: 1.5: example of PET image.

PET is utilized for diagnosing and evaluating a range of medical conditions. In oncology, PET is particularly valuable for detecting cancer, as it excels in identifying malignant tumors, staging the disease, monitoring treatment effectiveness, and detecting metastasis. Its ability to highlight areas of high metabolic activity allows PET scans to reveal tumors that may not be visible through other imaging methods like CT or MRI [Hegi-Johnson et al., 2022]. In neurology, PET is employed to study brain function and diagnose neurological disorders such as Alzheimer's disease, Parkinson's disease, and epilepsy. By assessing glucose metabolism in the brain, PET can detect changes associated with neurodegenerative conditions. In cardiology, PET imaging is used to evaluate myocardial perfusion, identify ischemic heart tissue, and determine the viability of heart muscle following a heart attack, making it a critical tool for cardiac assessment.

2.4.3. Comparison between SEPCT and PET

SPECT and PET are both nuclear imaging techniques used to diagnose and evaluate various medical conditions, but they differ in several key aspects:

1. Radiotracers and Imaging Mechanism:

   In nuclear medicine, SPECT and PET are both advanced imaging techniques used to create detailed 3D images of the body. SPECT utilizes gamma-ray-emitting radiotracers, such as technetium-99m, which release single photons. A gamma camera rotates around the patient to capture these emissions from various angles, reconstructing a three-dimensional image. On the other hand, PET employs positron-emitting radiotracers, like fluorodeoxyglucose, which release positrons. When these positrons collide with electrons, they annihilate and produce gamma rays. These gamma rays are then detected by the PET scanner, generating high-resolution 3D images that are particularly useful for assessing metabolic activity and detecting diseases such as cancer. While both techniques rely on radioactive tracers and gamma-ray detection, PET generally offers higher resolution and is

more sensitive to metabolic changes, whereas SPECT is more widely available and cost-effective.

2. Spatial Resolution:

SPECT and PET differ in spatial resolution, affecting diagnostic precision. SPECT has lower resolution (8-10 mm) due to gamma camera limitations, making it less sensitive for small lesions. PET offers higher resolution (4-6 mm) due to advanced detection of positron-emitting tracers, enabling detection of smaller abnormalities. This makes PET ideal for oncology and early detection, while SPECT remains widely accessible. PET's superior resolution often makes it the preferred choice for high-precision imaging.

3. Metabolic and Functional Imaging:

Based on their functional imaging capabilities, SPECT and PET serve different diagnostic roles. SPECT measures blood flow and tissue perfusion, making it ideal for cardiac studies such as myocardial perfusion imaging and brain perfusion scans. PET focuses on metabolic activity, like glucose uptake, excelling in oncology (cancer detection) and neurology (Alzheimer's studies). While SPECT visualizes physiological processes, PET tracks metabolic changes, offering critical insights into disease activity and progression.

4. Radiation Exposure:

SPECT and PET differ in radiation exposure due to their radiotracers. SPECT uses tracers like technetium-99m, which have longer half-lives and emit less energy, resulting in lower radiation doses and making it relatively safer. PET, however, uses positron-emitting tracers like FDG, which have shorter half-lives and higher energy emissions, leading to greater radiation exposure. While PET offers superior imaging for metabolic studies, its higher radiation dose is a key consideration, especially for repeated scans. SPECT generally poses a lower radiation risk compared to PET.

5. Cost and Availability:

SPECT is more widely available and cost-effective, as its scanners are common and radiotracers like technetium-99m are cheaper and longer-lasting, making it practical for many clinics. PET, however, is more expensive and less accessible due to costly scanners and the need for on-site cyclotrons to produce short-lived tracers like FDG. While PET offers advanced imaging for metabolic and oncological studies, its higher cost and infrastructure requirements limit its use compared to SPECT.

6. Clinical Applications:

SPECT and PET are used in different medical fields based on their imaging strengths. SPECT is commonly used in cardiology for heart function, neurology for brain perfusion, and bone scans for fractures or infections. PET is primarily used in

oncology for cancer detection and staging, neurology for diagnosing Alzheimer's and epilepsy, and cardiology for myocardial viability. While SPECT excels in structural and perfusion imaging, PET's focus on metabolic activity makes it ideal for functional and molecular diagnostics, especially in cancer and neurodegenerative diseases.

2.4.4. Advantages and Limitations of Nuclear Medicine

Nuclear medicine offers several advantages and limitations that are important to consider in clinical practice. One of its key strengths is its ability to provide functional and metabolic imaging, enabling the detection of disease processes at an early stage, often before structural changes are visible on other imaging modalities like CT or MRI. This makes it particularly valuable for early detection of conditions such as cancer, neurological disorders, and heart disease, allowing for timely intervention and treatment. Additionally, techniques like PET enable whole-body imaging, offering a comprehensive view of disease spread and metabolic activity, which is crucial for staging and treatment planning.

However, nuclear medicine also has its limitations. Radiation exposure, though generally low, is a concern, particularly for vulnerable populations such as pregnant women. The cost of certain procedures, especially PET scans, can be high, and these advanced imaging techniques may not be as widely accessible as CT or MRI. Furthermore, the availability of radiopharmaceuticals is a challenge, as these specialized agents often have short shelf lives and require specific production facilities, limiting their accessibility in some regions. Despite these limitations, nuclear medicine remains a powerful tool for diagnosing and managing a wide range of diseases.

5. Conclusion

The evolution of medical imaging, from basic X-rays to advanced modalities like MRI and PET, has revolutionized diagnostics and treatment in modern medicine. These imaging techniques provide detailed insights into the human body, enabling accurate diagnosis, effective treatment planning, and precise monitoring of disease progression. However, the complexity and volume of data generated by these modalities present significant challenges, particularly in extracting meaningful information from images. This is where medical image segmentation becomes crucial.

For instance, in MRI, segmentation can delineate brain tumors for surgical planning, while in PET scans, it can help quantify metabolic activity in cancer cells. Similarly, in X-rays and CT scans, segmentation aids in detecting fractures, infections, or other anomalies. As medical imaging continues to advance, the demand for accurate and efficient segmentation techniques grows, particularly with the increasing complexity of imaging data.

Medical imaging will continue to play a crucial role in modern medical practice, providing valuable information for diagnosis, treatment planning, and monitoring treatment response. Technological advancements will keep driving innovation in this field, offering new opportunities to enhance patient health and well-being.

The clinical applications of medical imaging are vast and include diagnosis, treatment planning, and monitoring treatment response. Future trends in medical imaging involve the increased use of artificial intelligence and machine learning to improve medical image analysis and interpretation, as well as the integration of multimodal data for a more comprehensive assessment of patient health.

Chapter 2 : Medical Image Segmentation

1. Introduction

Medical image segmentation is a fundamental process in the field of medical imaging, playing a pivotal role in diagnostics, treatment planning, and disease monitoring [Narayan et al., 2023]. At its core, segmentation involves partitioning an image into meaningful regions or structures, such as organs, tissues, or pathological areas, to enable detailed analysis and interpretation. The accuracy and efficiency of segmentation directly impact the quality of patient care, making it a critical component of modern medicine.

The evolution of medical imaging, from the simplicity of X-rays to the sophistication of modalities like MRI, CT, and PET, has significantly enhanced the ability to visualize and understand the human body. However, these advancements have also introduced new challenges. Medical images are often complex, with high variability in resolution, contrast, and noise levels. Additionally, anatomical structures can be intricate and overlapping, making it difficult to accurately identify and segment regions of interest.

Over the years, numerous medical image segmentation methods have been proposed, ranging from traditional thresholding and region-growing approaches to more recent machine learning and artificial intelligence-based techniques. In this chapter, we will review the various medical image segmentation methods, their advantages and limitations, as well as the challenges and opportunities associated with this critical task.

2. Definition of image segmentation

Image segmentation is a process in computer vision and image processing that involves partitioning a digital image into multiple segments or regions, each of which corresponds to different objects or parts of the image. The goal of image segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. This is typically achieved by assigning a label to every pixel in the image such that pixels with the same label share certain characteristics, such as color, intensity, or texture [Yu et al., 2023].

The essential role of the image segmentation lies in its ability to provide a structured and meaningful representation of visual information, enabling computer systems to understand and interact with their visual environment in a more sophisticated manner.

By partitioning an image into coherent segments, image segmentation allows for the identification and differentiation of various elements present in a visual scene, such as objects, edges, and textures.

This precise segmentation is fundamental for many applications, including object recognition, pattern detection, video surveillance, autonomous navigation, computer-aided diagnostic medicine, and many more.

There are several approaches and techniques used in image segmentation. Each image segmentation technique involves a series of specific operations to process and analyze

images. Each method is suited to specific contexts and has distinct advantages and limitations. The choice of method often depends on the characteristics of the image, the requirements for accuracy and performance, as well as the constraints of real-time processing when applicable.

The most common methods are as follows:

- Threshold based methods
- Edge based methods
- Region based methods
- Supervised classification based methods
- Unsupervised classification (Clustering) based methods

2. Thresholding Methods

Thresholding is a fundamental technique in the field of image segmentation [Jardim et al., 2023]. Its main goal is to convert a grayscale image into a binary image, where each pixel is assigned one of two values, usually 0 or 1, corresponding to black or white, respectively (cf.fig. 2.1). In simpler terms, thresholding can be understood as:

$$S(x,y) = \begin{cases} 1, if\ I(x,y) \geq T \\ 0, if\ I(x,y) < T \end{cases} \quad (2.1)$$

where

- $I(x,y)$ is the pixel value of the grayscale image at position (x,y);
- $S(x,y)$ is the pixel value of the binary image at position (x,y);
- T is the chosen threshold.

(a) (b)

Figure 2.1: Thresholding-based segmentation, (a) original image, (b) segmented image.

In image processing, thresholding methods can be broadly categorized into two types: global and local thresholding. This classification depends on how the threshold value is determined and applied across the image.

2.1. Global Thresholding Methods

Global thresholding relies on selecting a single, fixed threshold value, which is carefully determined based on the uniform characteristics present across the entire image or derived from its overall histogram. This approach is generally used when there is a clear contrast in the grayscale distribution between the foreground (or region of interest) and the background [Senthilkumaran 2016]. Prominent global thresholding methods include Otsu's Method, Iterative Thresholding, Minimum Error Thresholding, and Entropy-based Thresholding.

2.1.1. Otsu technique

The Otsu thresholding technique identifies the optimal threshold by maximizing the variance between the gray levels of an object and its background [Goh 2018]. The process begins by converting the image into grayscale, which comprises 256 gray levels ranging from 0 (black) to 255 (white). The method seeks to determine a gray value threshold that effectively separates the background (higher gray values) from the foreground (lower gray values). The optimal threshold corresponds to the point where this variance is maximized, thereby improving the contrast between black and white in the resulting binarized image.

2.1.2. Iterative Thresholding

Iterative thresholding is a method for binarizing images by iteratively refining the threshold based on the mean gray-scale values of the foreground and background. The process begins with an initial threshold, T1, often set as the average gray value of the image. The image is then divided into two regions: pixels with values above T1 (background) and those below T1 (foreground). The mean gray values of these regions are computed to determine a new threshold, T2. This iterative process continues, updating T1 with T2 in each step, until the difference between successive thresholds falls below a predefined tolerance level, T0 (typically T0 = 0.5 or 1). At this point, T2 is considered the optimal threshold [Sujji et al., 2013].

Iterative thresholding offers several advantages, including its simplicity and ease of implementation, making it a computationally efficient method for binarizing images. It dynamically adapts the threshold based on the image's gray-level distribution, eliminating the need for prior knowledge of the image's histogram. This method is particularly effective for images with a clear separation between foreground and background, as it guarantees convergence to an optimal threshold within a predefined tolerance level. However, iterative thresholding also has notable limitations. Its performance is highly dependent on the initial threshold choice, which can lead to suboptimal results if poorly selected. The method assumes a bimodal intensity distribution, making it less suitable for images with complex or overlapping intensity profiles. Additionally, it is sensitive to noise and artifacts, which can distort the mean gray values and result in inaccurate thresholds. Computational costs can also increase for high-resolution images, especially when a small convergence tolerance is used. Furthermore, the technique is primarily designed for grayscale images and may struggle with non-uniform illumination or color images without preprocessing. Despite these drawbacks,

iterative thresholding remains a valuable tool for applications where simplicity and adaptability are prioritized, provided the image characteristics align with its assumptions.

2.1.3. Minimum Error Thresholding

The Minimum Error Thresholding technique is based on the assumption that the gray values of pixels in an image's foreground and background follow a normal distribution. The goal of this method is to determine an optimal threshold T that reduces the overall classification error during segmentation. This error is quantified by analyzing the probability density functions of each segment, taking into account their respective probabilities and variances.

2.1.4. The Entropy Method

The entropy method works by calculating the entropy of the image for different brightness thresholds and selecting the brightness threshold that maximizes the entropy. An image with high entropy contains a lot of information, while an image with low entropy contains little information. The brightness threshold that maximizes the entropy is chosen as the optimal threshold [Yin, 2002].

This method offers several advantages, including its effectiveness for images with complex brightness distributions, its robustness to noise and lighting variations, and its adaptability to different types of images. However, it also has some drawbacks, such as its computational complexity and sensitivity to the choice of entropy calculation method. These factors should be carefully considered when applying this technique in image processing tasks.

2.2. Local Thresholding Methods

These methods determine threshold values based on the statistical features of each distinct neighborhood, taking into account factors like brightness, contrast, and textural details. As a result, each pixel in the image is classified according to the attributes of its neighboring pixels. They require the use of more advanced algorithms that carefully analyze the local properties of each segment in the image. Various local adaptive thresholding techniques include Niblack’s Method, Sauvola’s Method, and Bernsen’s Method [Saxena et al., 2019]

2.2.1. Niblack’s Method

Niblack’s technique is a prominent approach in local adaptive thresholding. It computes the local mean (m) and standard deviation (s) of pixel values within a defined window centered on each pixel. By utilizing the mean to evaluate local brightness and the standard deviation to measure contrast or texture, this method dynamically determines thresholds, enabling efficient image segmentation.

2.2.2. Sauvola’s Method

This method faces challenges with low-texture backgrounds, where subtle details might exceed the established threshold. Sauvola improved this technique to more effectively

manage diverse backgrounds and lighting conditions by integrating the dynamic range of the standard deviation into the threshold calculation. This adjustment enables the method to adaptively enhance the dynamic range of the standard deviation, improving its performance.

2.2.3. Bernsen's Method

Bernsen's technique emphasizes adaptive segmentation by utilizing the local contrast of pixel regions to distinguish between high-contrast areas which is often associated with edges or text and low-contrast regions, which typically represent uniform backgrounds. This method is particularly effective at identifying subtle variations in images and accentuating important features against varied backgrounds [Senthilkumaran & Vaithegi, 2016].

3. Edge-Based Segmentation

An edge represents a boundary between two homogeneous regions and is characterized by a local variation in image intensity. Its detection involves identifying and pinpointing sharp discontinuities within an image. In edge-based segmentation techniques, the process begins by detecting the contours of objects and the boundaries separating objects from the background. These edges are then connected to form complete object boundaries, enabling the segmentation of the desired regions. Discontinuity-based segmentation methods are particularly effective at identifying abrupt changes in intensity values. The core principle of edge detection lies in locating areas where significant changes in image characteristics occur [Sharma et al., 2013].

Edge-based segmentation offers several advantages. It is highly precise, capable of segmenting complex objects with irregular contours, and robust, as it is less sensitive to noise and lighting variations. Additionally, it is relatively simple to understand and implement. However, this method also has some limitations. It can be sensitive to incomplete or noisy edges, and it may struggle to effectively segment overlapping objects. These factors should be considered when applying edge-based segmentation in image processing tasks.

Gradient-based edge detection is a simple and effective method for detecting edges in an image. It is a spatial filter that calculates the gradient of an image's intensity. The gradient measures the variation in intensity across a given direction in the image. It is computed using two filters: one for the horizontal gradient and one for the vertical gradient. These filters are weight matrices that are multiplied by the pixels of the image. The result is a set of values representing the intensity gradient in each direction. The gradient values are then used to detect edges. Edges are typically identified as points where the gradient is high. These points are connected by lines to form the edges of the image (cf.fig.2.2).

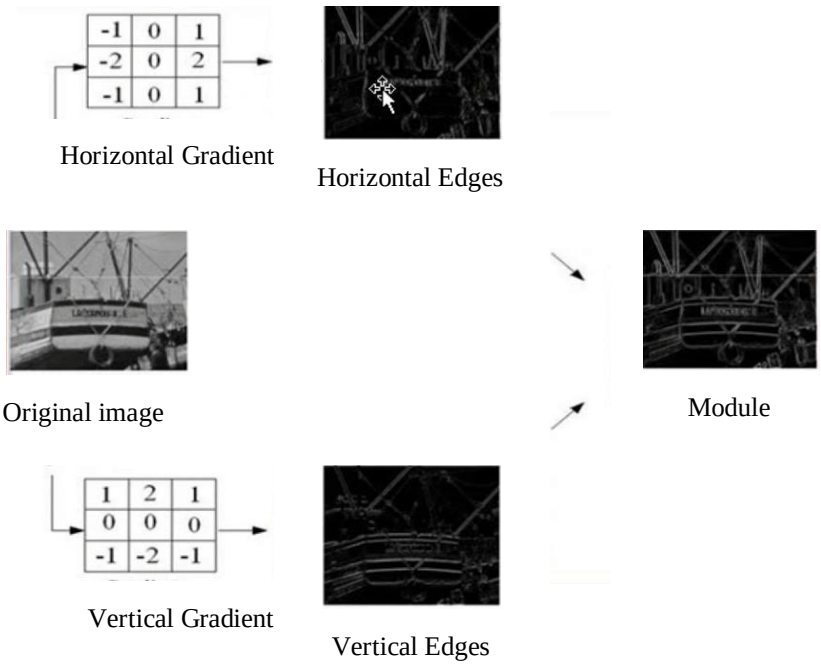


Figure 2.2: contour detection by gradient approach

Edge detection plays a vital role in edge-based segmentation. In images, edges are defined by two key attributes: direction and magnitude. Along the edge's direction, pixel values tend to change gradually, whereas perpendicular to the edge, they exhibit sharp transitions. Because of these properties, first and second-order derivatives are widely employed to identify and characterize edges effectively.

3.1. First-order differential operators

First-order differential operators are fundamental tools in edge detection, as they identify edges by computing intensity gradients across an image. Among the most widely used operators are the Roberts Cross Operator, Prewitt Operator, Sobel Operator, and Canny Edge Detector. Each of these techniques functions by approximating the first derivative of pixel intensities, emphasizing regions where abrupt changes in brightness occurrence indicating potential edges [Acharjya et al., 2012].

- The Roberts Cross Operator employs a simple 2x2 kernel to detect edges at 45-degree angles, making it computationally efficient but sensitive to noise.
- The Prewitt and Sobel Operators use 3x3 convolution kernels to estimate horizontal and vertical gradients, with the Sobel operator incorporating weighted smoothing for better noise resistance.
- The Canny Edge Detector is a more sophisticated approach, combining Gaussian smoothing, gradient computation, non-maximum suppression, and hysteresis thresholding to produce high-precision edge maps with minimal noise interference.

These operators vary in complexity and robustness, making them suitable for different applications depending on accuracy and computational efficiency requirements.

3.2. Second-order differential operators

Second-order differential operators play a critical role in edge detection by identifying intensity discontinuities based on curvature and gradient changes. Two prominent examples are the Laplacian Operator and the Laplacian of Gaussian (LoG). Unlike first-order operators that detect edges by locating gradient maxima, second-order operators rely on zero-crossings in the second derivative, which correspond to sharp intensity transitions.

- The Laplacian Operator applies a second-derivative-based convolution kernel to highlight rapid changes in pixel intensity, making it highly sensitive to noise.
- The Laplacian of Gaussian (LoG) improves robustness by first smoothing the image with a Gaussian filter to reduce noise before applying the Laplacian, resulting in more accurate edge localization.

These operators function by convolving a predefined template matrix (kernel) with the image's pixel value matrix, effectively computing local gradients or curvatures at each pixel. While second-order methods excel at detecting fine edges and corners, their sensitivity to noise often necessitates preprocessing steps, such as Gaussian smoothing in the case of LoG [Veelaert & Teelen, 2009].

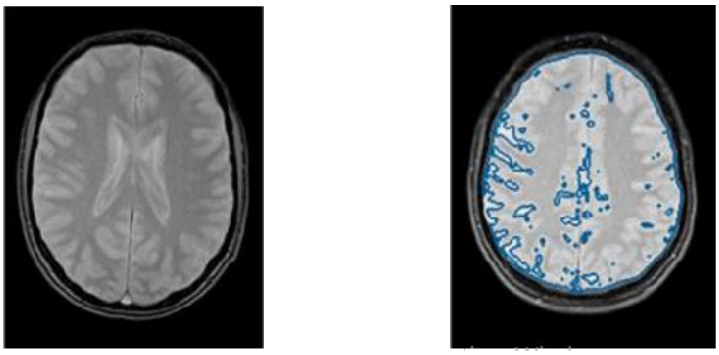
3.3. Operator's characteristics

Edge detection operators exhibit distinct trade-offs between performance and computational efficiency:

- The Roberts Cross offers simplicity but suffers from noise sensitivity;
- Prewitt maintains directional sensitivity while remaining vulnerable to noise;
- Sobel improves noise suppression at the cost of edge blurring;
- Canny delivers superior accuracy through multi-stage processing but requires careful parameter tuning;
- Laplacian precisely localizes edge centers yet amplifies noise without directional information;
- Laplacian of Gaussian (LoG) combines Gaussian smoothing with second-order differentiation for balanced noise robustness and localization, albeit with increased computational overhead and potential loss of fine details.

4. Region Based Segmentation

Region based segmentation is a technique that partitions an image into meaningful regions based on pixel similarity, such as intensity, color, texture, or other statistical properties. Unlike edge-based segmentation, which detects boundaries, region-based methods group pixels into coherent regions by analyzing their homogeneity. Figure bellow presents an example of medical image segmented with this technique.



(a) (b)

Figure 2.3: Region based segmentation, (a) original image, (b) segmented image.

While region-based segmentation is simple to implement, robust to lighting changes, and effective for objects of diverse sizes/shapes, its performance declines with complex images (overlapping objects, irregular contours, or noisy data), potentially yielding flawed results [Karthick et al., 2014].

Region-based segmentation primarily employs three core methodologies to partition digital images into coherent regions by leveraging pixel similarity criteria:

1. Region Growing.
2. Split and Merge.
3. Watershed lines.

4.1. Region Growing

Region growing is an image segmentation method that works by grouping pixels into regions based on their similarity. It starts with a set of seed points, pixels known to belong to a specific region, and then expands these seeds to include neighboring pixels that meet similarity criteria. This process continues until all seeds are fully expanded and every pixel in the image is assigned to a region [Shrivastava & Bharti, 2020].

Similarity between pixels can be measured using various metrics, such as Euclidean distance or Manhattan distance. The choice of metric depends on the image characteristics and segmentation goals.

Region growing can be applied to both grayscale and color images. For color images, the method may be implemented by processing each color channel independently.

Region growing offers advantages such as simplicity, effectiveness, robustness to noise and lighting variations, and adaptability to different image types; however, its performance heavily depends on the initial seed selection, potentially leading to poorly segmented regions if the seeds are not optimally chosen.

4.2. Split-and-Merge

The split-and-merge algorithm takes an opposite approach to region growing by beginning with the entire image as a single region. The process then iteratively splits any regions that violate a homogeneity criterion while simultaneously merging adjacent regions that demonstrate sufficient similarity. This dual-phase methodology has become a fundamental technique with broad applications across multiple domains.

In image segmentation implementations, the homogeneity criteria typically mirror those used in region growing approaches. The splitting operation conventionally divides non-homogeneous regions into four equal rectangular partitions, while the merging phase selectively combines neighboring regions based on carefully defined similarity measures - a crucial aspect determining the algorithm's effectiveness. A key constraint requires that only spatially adjacent regions may merge. The algorithm converges when the splitting phase can no longer generate new regions, indicating complete segmentation.

The split-and-merge algorithm offers significant advantages, including computational efficiency, robustness, and adaptability to various image types, while consistently producing well-segmented regions; however, it presents notable drawbacks such as high computational complexity and sensitivity to the selection of splitting and merging parameters, which can critically impact segmentation quality [Zaitoun & Aqel, 2015].

4.3. Watershed Lines

Watershed segmentation is a popular image processing technique inspired by topographic flooding, where pixel intensities are treated as elevation levels. The method works by flooding the image's gradient magnitude from regional minima, creating boundaries (watershed lines) where different "catchment basins" meet. This approach is particularly effective for separating touching or overlapping objects in images, making it valuable in medical imaging, material science, and biological analysis. Watershed transformation excels at detecting precise edges and works well with both grayscale and color images [Mohanapriya & Kalaavathi, 2019].

The primary advantages of watershed segmentation include its ability to detect closed boundaries and delineate objects with high precision, even when they are in close contact. It is also conceptually intuitive and works without prior knowledge of the number of objects in the image. However, drawbacks include sensitivity to noise, leading to excessive segmentation, and high computational cost for large images. Pre-processing steps like smoothing or marker-controlled watershed (using predefined seed points) are often required to improve results. Additionally, the method may struggle with low-contrast images where gradient differences are insufficient for proper boundary detection. Despite these limitations, watershed remains a powerful tool when combined with appropriate pre- and post-processing techniques.

5. Classification Methods for Segmentation

Image segmentation can be effectively approached using classification-based methods, which involve assigning labels to individual pixels or regions within an image based on their underlying features such as color, texture, intensity, or spatial context. These features are used to group similar pixels together, thereby delineating meaningful structures or objects within the image.

Classification-based segmentation techniques can be broadly categorized into supervised and unsupervised methods. Supervised approaches rely on labeled training data, where examples of correctly segmented images are used to train a model to recognize similar patterns in new, unseen images. These methods often employ machine learning algorithms such as support vector machines (SVM), decision trees, or deep learning networks like convolutional neural networks (CNNs). In contrast, unsupervised methods, known also as clustering methods do not require labeled data and instead aim to discover inherent structures within the image by clustering similar pixels together based on feature similarity. Common unsupervised techniques include k-means clustering, Gaussian mixture models, and hierarchical clustering. Each approach has its strengths and trade-offs, with supervised methods typically achieving higher accuracy given quality training data, while unsupervised methods offer greater flexibility and are particularly useful in scenarios where labeled data is scarce or unavailable.

5.1. Supervised Classification Methods for Segmentation

Supervised classification methods for image segmentation rely on labeled training data to train a model that can classify individual pixels or regions into predefined categories. These approaches follow a structured pipeline beginning with (1) feature extraction, where key characteristics such as pixel intensity, texture patterns, edge information, and spatial relationships are quantified to represent each pixel or region. (2) These extracted features are then used to train a machine learning classifier such as k-Nearest Neighbors (K-NN), Random Forests (RF), or Support Vector Machines (SVM) on annotated datasets, where each pixel or region is associated with a ground truth label. (3) Once trained, the model can predict labels for new, unseen image data by analyzing their features and assigning them to the most probable category.

With the advent of deep learning, supervised segmentation has shifted toward end-to-end trainable models such as INet [Weng & Zhu, 2021], FCNs [Jian et al., 2018], which automate feature extraction through convolutional layers and achieve state-of-the-art accuracy [Liu et al., 2021a].

5.1.1. K-Nearest Neighbors (K-NN)

K-NN [Cunningham & Delany, 2021] is a simple, non-parametric supervised learning algorithm used for pixel-wise classification in image segmentation. It relies on feature similarity to assign labels based on the closest examples in the training data.

The K-Nearest Neighbors (K-NN) algorithm follows a straightforward yet effective workflow for image segmentation. First, feature extraction transforms each pixel into a numerical representation, typically a feature vector that may include attributes such as intensity values, texture descriptors (Haralick features for medical images), color channels (like RGB), or spatial coordinates to incorporate positional context. Unlike parametric models, K-NN adopts a lazy learning approach during the training phase, where it simply stores all labeled feature vectors along with their ground truth classifications in memory without deriving an explicit model. During the prediction phase, the algorithm processes a new pixel by calculating its distance (Euclidean or Manhattan for examples) to every labeled pixel in the training set, identifies the K closest neighbors, and assigns the majority class label among them to the target pixel. While computationally intensive this method's simplicity and lack of assumptions about data distribution make it a versatile baseline for segmentation tasks, particularly in scenarios with limited training data or low-dimensional feature spaces.

5.1.2. Random Forest

Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes of the individual trees. In the context of image segmentation, Random Forest classifiers are used to assign a label (like object or background) to each pixel or image patch based on its features [Parmar et al., 2018].

In segmentation tasks, features might include color intensity, texture, edge information, spatial location, or filter responses. These features are extracted for each pixel (or region) and used as input to the Random Forest. The classifier learns from labeled training data and generalizes to segment unseen images by classifying each pixel into one of the predefined classes.

This method follows the pipeline below

1. Feature Extraction: Extract relevant features per pixel or region.
2. Training: Use labeled examples to train the Random Forest.
3. Prediction: Classify each pixel in a new image based on learned decision trees.

5.1.3. Support Vector Machines (SVM)

SVM [Jasti et al., 2022] is a powerful supervised learning algorithm commonly used for classification tasks. In image segmentation, SVMs are used to assign labels to individual pixels or regions of an image by learning from feature representations derived from training data.

An SVM works by finding the optimal hyperplane that separates data points of different classes with the maximum margin. When data are not linearly separable, kernel functions like radial basis function (RBF) or polynomial are employed to map data into higher-dimensional spaces where separation becomes feasible. Each pixel or region in the image is represented as a feature vector. Common features include:

- Color
- Texture

- Edges
- Position
- Neighborhood statistics: Mean, variance in a local window

The training process follows the steps below

1. Annotate a set of training images with class labels (pixel-level or region-level).
2. Extract features for each labeled pixel.
3. Train the SVM using these features.
4. For new (unlabeled) images, extract the same features and use the trained SVM to classify each pixel.

SVMs offer several advantages that make them particularly well-suited for image segmentation tasks. They are highly effective in high-dimensional spaces, which are beneficial when working with rich pixel-wise feature sets that include color, texture, and spatial information. SVMs are also robust in scenarios where there is a clear margin between classes, often resulting in high classification accuracy. Their flexibility is enhanced by the use of kernel functions, which enable the algorithm to model complex, non-linear decision boundaries by projecting data into higher-dimensional feature spaces. Furthermore, SVMs perform well even with relatively small datasets, making them an attractive option when annotated training data are limited.

Machine learning approaches in this domain are particularly useful when interpretability and computational efficiency are prioritized, but they often require careful feature engineering to achieve robust performance. However, a key limitation of these methods is their dependency on high-quality labeled data and their potential struggle with complex, heterogeneous structures where manually designed features may not capture sufficient discriminative information.

Nevertheless, these approaches remain relevant in scenarios with limited training data or constrained computational resources, offering a balance between performance and simplicity.

5.1.4. Convolutional neural network (CNN)

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed to process grid-like data, such as images. CNNs have become the dominant approach for image segmentation tasks due to their ability to automatically learn hierarchical features directly from raw pixel data and their effectiveness in capturing spatial context.

In image segmentation, CNNs classify each pixel (or group of pixels) in an image into a specific category, enabling precise delineation of objects or regions such as tumors, roads, or people.

A typical CNN for segmentation consists of:

- Convolutional layers: Learn local patterns such as edges, textures, or object parts.

- Pooling layers: Reduce spatial resolution to capture more abstract representations.
- Fully connected layers (in traditional CNNs): Used for classification, though often omitted in modern segmentation networks.
- Upsampling/Deconvolution layers: Restore spatial resolution for pixel-wise prediction.

Key Architectures in CNN-Based Segmentation are:

- Fully Convolutional Networks (FCN): The major CNN architecture designed for segmentation [Jian et al., 2018].
- U-Net: Popular in biomedical image segmentation; uses an encoder-decoder structure with skip connections to preserve spatial detail [Ronneberger et al., 2015].
- SegNet: Employs encoder-decoder structure with index-based upsampling to improve memory efficiency [Badrinarayanan et al., 2017].
- DeepLab: Uses atrous (dilated) convolutions and Conditional Random Fields (CRFs) to improve boundary accuracy [Chen et al., 2017].
- Mask R-CNN: Extends object detection by adding a branch for pixel-level object masks [He et al., 2017].

5.2. Clustering methods

Clustering-based methods is a popular unsupervised approach that groups pixels into clusters based on similarity in features such as color, intensity, or texture. Unlike supervised methods, clustering does not require labeled training data, making it widely applicable in scenarios where manual annotation is impractical [Mokhtari & Debakla, 2018] [Saxena et al., 2019].

Clustering algorithms categorize pixels into groups (clusters) where intra-cluster similarity is high and inter-cluster similarity is low. Common features used for clustering include color, intensity, spatial coordinates and texture (cf.fig.2.4).

Clustering methods are classified into two distinct groups: hierarchical and partitional techniques.

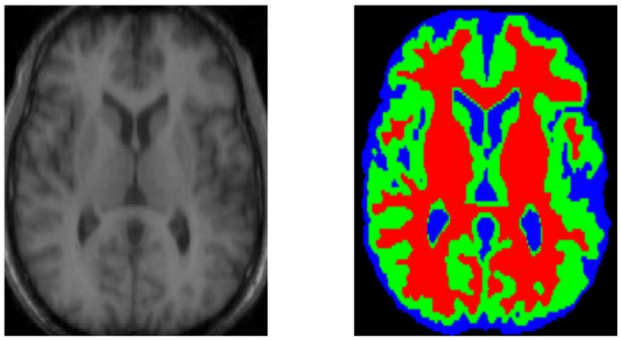


Figure 2.4: Brain MRI image segmented in three clusters.

5.2.1. Hierarchical clustering

Hierarchical clustering organizes data into clusters through an iterative process, using either a bottom-up (agglomerative) or top-down (divisive) approach. These methods build a dendrogram -a binary tree structure (cf.fig.2.5) - that visually represents the nested grouping of patterns. Hierarchical clustering is broadly classified into two types: agglomerative, which merges smaller clusters into larger ones, and divisive, which recursively splits larger clusters into finer subgroups.

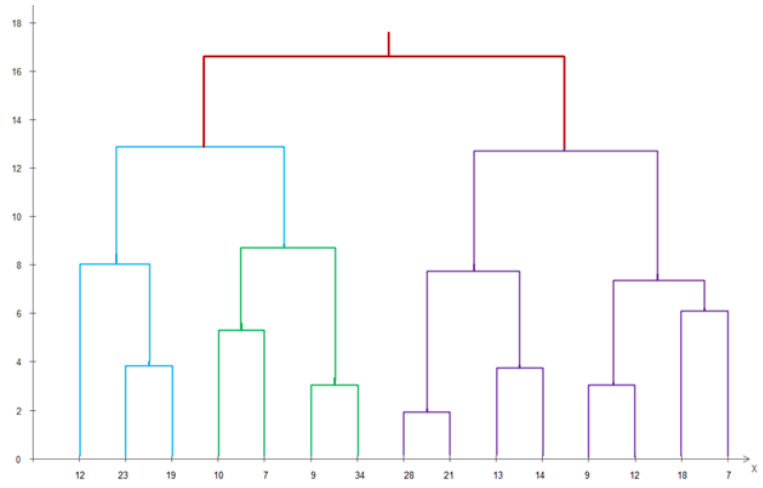


Figure 2.5: Dendogram.

An agglomerative hierarchical clustering start with each data point as its own cluster and iteratively merges the most similar pairs, building the hierarchy bottom-up. In contrast, the divisive approach begins with all data points in a single cluster and recursively splits them top-down into smaller subgroups, continuing until each point is isolated or a stopping condition is met.

Hierarchical clustering offers several benefits, including the ability to reveal nested cluster structures through dendrograms, which provide intuitive visualizations of data relationships. It does not require pre-specifying the number of clusters, making it useful for exploratory analysis. Additionally, it can handle arbitrary cluster shapes and is applicable to various data types, given a suitable similarity measure.

However, hierarchical clustering has notable limitations. It is computationally expensive, with a time complexity of $O(n^3)$ for agglomerative methods, making it impractical for large datasets. The approach is also sensitive to noise and outliers, which can distort the hierarchy. Furthermore, since decisions on merging or splitting clusters are greedy and irreversible, early errors can propagate, leading to suboptimal results. Finally, interpreting dendrograms can be subjective, as the choice of where to "cut" the hierarchy to define clusters is often arbitrary [Mittal et al., 2021].

5.2.2. Partitional Clustering

Partitional clustering methods group data points into non-overlapping clusters by optimizing an objective function, maximizing intra-cluster similarity while minimizing inter-cluster similarity. Typically, similarity is measured using metrics like Euclidean distance, and the algorithm iteratively refines clusters to minimize within-cluster variance.

A key strength of partitioning algorithms is their iterative refinement of clustering quality. This capability is absent in hierarchical clustering methods. However, these algorithms suffer from notable limitations. They require a predefined number of clusters, often resulting in inadequate cluster descriptors. Additionally, their performance is highly sensitive to initialization and can be severely compromised by noise and outliers. Furthermore, they struggle with clusters of uneven sizes, varying densities, or non-convex geometries [Ikotun et al., 2023].

Partitional clustering includes hard clustering and soft clustering exemplified by Fuzzy c-means clustering (FCM).

5.2.2.1. Hard clustering

Hard clustering (also called crisp or exclusive clustering) is a partitioning method where each data point $x_i \in X$ (X presents the all data) is definitively assigned to exactly one cluster $C_j \in \{C_1,...,C_k\}$, such that:

- Membership function $\mu(x_i,C_j) \in \{0,1\}$ (binary assignment)
- $\cup C_j = X$ (complete coverage)
- $C_j \cap C_k = \emptyset \ \forall \ j \neq k$ (mutual exclusivity)

This creates strictly delineated cluster boundaries. Common hard clustering algorithms include:

K-Means:

The k-means algorithm is a simple and efficient clustering algorithm often used for image segmentation and object detection. The k-means algorithm works in several steps:

1. Initialization: Select k initial centroids (cluster centers), either randomly or using a specific method.
2. Assignment: Assign each data point to the nearest centroid based on distance (usually Euclidean distance).
3. Update: Recalculate the centroids by taking the mean of all points assigned to each cluster.
4. Iteration: Repeat the assignment and update steps until convergence (when centroids no longer change significantly).

The k-means algorithm offers several advantages: it is simple to implement, efficient for image segmentation and object detection, and robust to noise and outliers. However, it also has limitations: it is sensitive to the initial choice of centroids, may not converge to an optimal solution, and cannot detect clusters with complex shapes [Fränti & Sieranoja, 2019].

Mean shift:

Mean Shift Clustering is a non-parametric, density-based clustering algorithm particularly effective for tasks like image segmentation. This method does not require predefining the number of clusters. Instead, it operates by:

1. Sliding a Kernel Window: A window (Gaussian kernel) moves across the data space, calculating the mean of data points within its bandwidth.
2. Shifting to Higher Density: The window iteratively shifts toward regions of maximum density (gradient ascent) until convergence.
3. Cluster Formation: Points converging to the same mode (peak density) are grouped into a cluster.

Mean Shift clustering is ideal for image segmentation because it automatically finds clusters without predefined numbers, handles noise well, and creates smooth segments by combining color and spatial data. Its density-based approach preserves object boundaries naturally, making it perfect for complex images like medical scans or satellite photos [Wang et al., 2017].

DBSCAN:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm that groups data points into clusters based on their spatial density, while identifying noise points that do not belong to any cluster. DBSCAN does not require predefining the number of clusters and can detect arbitrarily shaped clusters, making it particularly useful for datasets with irregular geometries. The algorithm operates by defining a neighborhood around each point with a radius ε and requiring a minimum number of points (min_samples) within this neighborhood to form a dense region. Points are classified as *core points* (dense regions), *border points* (on the edges of dense regions), or *noise points* (isolated outliers) [Schubert et al., 2017]. DBSCAN follows the steps bellow:

Step 1: Initialization
    Mark all points as unvisited.
    Initialize an empty list of clusters.
Step 2: Core Point Detection
    For each unvisited point *p:*
        Find all points in its *ε-neighborhood.*
    If the neighborhood has ≥ min_samples points:
        Mark *p* as a *core point.*
        Create a new cluster.
        Expand the cluster using density reachability.
    Else, mark *p* as noise (temporarily).
Step 3: Cluster Expansion (Density Reachability)
    For each core point *p* in the current cluster:
        Find all *ε-neighbors* of *p*.
        For each neighbor *q*:

If q is unvisited: Mark as visited.

If q has ≥ min_samples neighbors, it is a core point → add its neighbors to the cluster.

If q is not yet assigned to any cluster, add it to the current cluster.

Step 4: Repeat Until All Points Are Processed

Continue until all points are either assigned to a cluster or marked as noise.

DBSCAN excels in handling noise and outliers, as it explicitly identifies and excludes them from clusters. However, its performance is sensitive to the choice of ε and *min_samples*, and it struggles with datasets where clusters have varying densities.

5.2.2.2. Soft clustering

Soft clustering assigns data points to clusters probabilistically, allowing for partial membership in multiple clusters. Unlike hard clustering (k-means), where each point belongs to only one cluster, soft clustering captures uncertainty and overlapping structures in data. Soft clustering algorithms include:

Gaussian Mixture Model (GMM) Algorithm:

GMM is a probabilistic soft-clustering method that models data as a mixture of K Gaussian distributions (K clusters).

The Gaussian mixture model assigns a probability to each data point x of belonging to a cluster. The probability of data point coming from Gaussian cluster i is expressed as:

$$p(x) = \sum_{i=1}^{K} \pi_i N(x|\mu_i, \Sigma_i) \tag{2.2}$$

where:

π_i = mixing coefficient (weight) for cluster i
μ_i = mean vector of cluster i
Σ_i = covariance matrix of cluster i

It uses the Expectation-Maximization (EM) algorithm to estimate cluster parameters.

1. Expectation Step: In this step, the algorithm calculates the probability that each data point belongs to each cluster based on the current parameter estimates (mean, covariance, mixing coefficients).
2. Maximization Step: After estimating the probabilities, the algorithm updates the parameters (mean, covariance, and mixing coefficients) to better fit the data.

These two steps are repeated until the model converges, meaning the parameters no longer change significantly between iterations.

GMMs offer probabilistic soft clustering, allowing data points to belong to multiple clusters simultaneously through membership probabilities, making them ideal for overlapping datasets. GMMs model flexible cluster shapes (spherical, elliptical, or tilted) via customizable covariance matrices, adapting to complex data distributions. As a generative model, GMMs provide uncertainty quantification, useful for confidence estimation in applications like anomaly detection. However, they exhibit sensitivity to initialization, often requiring multiple restarts for stable results.

A key limitation is their Gaussian assumption, which may underperform on non-normal data. Additionally, GMMs become computationally intensive in high-dimensional spaces due to covariance matrix inversions.

Fuzzy C-Means (FCM) Clustering:

FCM is a prominent soft clustering technique that extends traditional k-means by allowing partial membership of data points across multiple clusters [Bezdek, 1981]. Unlike hard clustering methods that assign each point to a single cluster, FCM employs a membership matrix $U = [u_{ij}]$ where $u_{ij} \in [0,1]$ represents the degree of belongingness of the i^{th} data point to the j^{th} cluster.

FCM algorithm iteratively optimizes an objective function J that incorporates weighted distances between data points and cluster centroids, with the weighting exponent m controlling the fuzziness of the resulting partitions.

$$J(U, C) = \sum_{i=1}^{K} \sum_{j=1}^{N} u_{ij}^{m} d^2(x_j, c_i) \tag{2.3}$$

U and $C = (c_1, c_2, \dots, c_K)$ are the memberships degrees matrix and a vector of clusters centers respectively. m ∈ [1, ∞[is to control fuzziness, $d^2(x_j, c_i)$ is the grayscale Euclidean distance and u_{ij} is the membership degree of the point j in the i^{th} cluster c_i

During execution, FCM alternates between calculating membership degrees based on current centroids and updating centroids according to current memberships, converging when either the change in centroids or objective function falls below a threshold.

This approach provides several advantages: it naturally handles overlapping cluster boundaries, offers more nuanced interpretation of ambiguous data points, and demonstrates greater robustness to noise compared to crisp clustering methods.

However, FCM requires careful selection of the fuzzifier parameter m and remains sensitive to initial centroid placement. The method finds particular utility in applications where cluster boundaries are inherently vague, such as medical image analysis, market segmentation, and bioinformatics, while its computational complexity remains comparable to traditional k-means clustering.

5. Conclusion

Medical image segmentation remains one of the most challenging yet crucial tasks in computational diagnostics. While traditional methods have laid important groundwork, they each suffer from fundamental limitations that restrict their clinical applicability:

1. Thresholding collapses when confronted with overlapping tissue intensities or inhomogeneous contrast enhancement, producing jagged, unrealistic boundaries that fail to capture pathological nuances.
2. Region-based methods bleed across anatomical borders, unable to distinguish true tissue interfaces from partial volume effects or imaging artifacts.
3. Edge detection disintegrates when faced with the low contrast-to-noise ratios characteristic of early-stage lesions or diffuse pathologies.
4. Rigid clustering methods (K-means, GMM) impose artificial binary decisions on inherently gradational biological transitions, discarding the probabilistic nature of medical interpretation.

These constraints underscore the urgent need for segmentation methods that preserve uncertainty in ambiguous regions, model the continuous nature of tissue interfaces, and adapt to variable imaging conditions ensuring robustness across diverse clinical and experimental scenarios

The solution emerges in the next chapter through Fuzzy C-Means (FCM) clustering, a paradigm-shifting approach that:

1. Shatters the binary segmentation fallacy through probabilistic membership functions
2. Captures transitional tissue states via smooth, overlapping cluster assignments
3. Mirrors radiologist reasoning by maintaining diagnostic uncertainty where appropriate
4. Provides tunable precision through its fuzzy parameter

The following chapter will dissect FCM's mathematical foundations, demonstrate its superiority in handling medical imaging ambiguities, and reveal how its soft decision boundaries enable more natural integration with downstream diagnostic AI systems and finally how it is enhanced through several contributions by modifying FCM’s objective function or optimizing its parameters using bio-inspired methods.

Chapter 3: FCM Optimization based on Bio-Inspired Methods

1. Introduction

Image segmentation is a critical task in computer vision and medical imaging, where the goal is to partition an image into meaningful regions for further analysis. Traditional segmentation methods often struggle with the inherent ambiguity, noise, and intensity inhomogeneity present in real-world images. Fuzzy C-Means (FCM) clustering has emerged as a powerful tool in this domain, offering a flexible approach by allowing pixels to belong to multiple clusters with varying degrees of membership. Unlike crisp (or hard) clustering methods like K-means, FCM's capability as soft segmentation makes it particularly effective for handling overlapping structures, such as tissues in medical images or objects with blurred boundaries in natural scenes.

The standard FCM algorithm minimizes an objective function that weighs pixel intensities against cluster centroids, making it suitable for intensity-based segmentation. However, conventional FCM has limitations, including sensitivity to noise, high computational cost, and dependence on initialization. To address these challenges, numerous FCM variants have been developed, specifically tailored for image segmentation:

- Spatial FCM (SFCM): Incorporates neighborhood pixel information to improve robustness against noise.

- Kernel FCM (KFCM) and Spatial KFCM (SKFCM): Use kernel functions to handle non-linear intensity distributions and to improve robustness against noise.

- Adaptive FCM (AFCM): Dynamically adjusts parameters based on local image statistics.

- Type-2 FCM (T2FCM): Enhances uncertainty modeling for low-contrast images.

Recent research has focused on hybridizing FCM with bio-inspired optimization algorithms such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) and other bio-inspired algorithms to optimize cluster initialization, improve convergence, and enhance segmentation accuracy. These hybridization approaches have shown significant promise in medical imaging (such as tumor detection), where precision and computational efficiency are paramount. This chapter explores the foundational FCM algorithm, its variants in image segmentation, and the growing impact of bio-inspired techniques in advancing fuzzy clustering for this complex task.

2. Fuzzy C-Means (FCM) for image segmentation

The FCM algorithms firstly introduced by Dunn [Dunn, 1974] and generalized by Bezdek [Bezdek, 1981] are a family of clustering algorithms based on a fuzzy objective function. They are considered as soft clustering in the way that each element of the data to be clustered may belong to more than one cluster with deferent degrees of membership. The objective function is optimized in an iterative way and at the end of the process; each element is assigned to the cluster in which it has the highest membership.

Let $I = (x_1, x_2, \dots, x_N)$ an image of N pixels to be clustered into K ($2 < K \ll N$) clusters, where x_i represents data features. The FCM objective function is formulated as [Bezdek, 1981]:

$$J(U,C) = \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_j, c_i) \tag{3.1}$$

U and $C = (c_1, c_2, \dots, c_K)$ are the memberships degrees matrix and a vector of clusters centers respectively. m ∈ [1, ∞[is to control fuzziness, $d^2(x_j, c_i)$ is the grayscale Euclidean distance and u_{ij} is the membership degree of the pixel j in the i^{th} cluster c_i which must check the following constraints:

$\forall i \in [1,K], j \in [1,N]$:

$$\sum_{i=1}^{K} u_{ij} = 1, \qquad u_{ij} \in [0,1], \qquad 0 \le \sum_{j=1}^{N} u_{ij} \le N \tag{3.2}$$

$J(U,C)$ is optimized, by introducing the Lagrange multipliers λ_i [Dunn, 1974] to incorporate the constraint in (2). This yields function $J(U,C,\lambda)$ to be minimized:

$$J(U,C,\lambda) = \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_j, c_i) + \sum_{j=1}^{N} \lambda_j \left[\sum_{i=1}^{k} u_{ij} - 1 \right] \tag{3.3}$$

J is alternately optimized in two steps:

Step 1: Optimizing the Membership Degrees:

U is an optimal value of J if $\frac{\partial J}{\partial u_{kl}} = 0$ and $\frac{\partial J}{\partial \lambda} = 0$ which lead to :

$$\frac{\partial J}{\partial \lambda_j} = \sum_{i=1}^{k} u_{ij} - 1 = 0 \tag{3.4}$$

$$\frac{\partial J}{\partial u_{ij}} = m u_{ij}^{m-1} d^2(x_j, c_i) + \lambda_j = 0 \tag{3.5}$$

Where $i = 1,2,\dots,K$ and $j = 1,2,\dots,N$.

From Equation (3.5), we obtain:

$$u_{ij} = \left(\frac{-\lambda_j}{m d^2(x_j, c_i)} \right)^{\frac{1}{m-1}} \tag{3.6}$$

By combining Equation (3.4) and (3.6), we get:

$$\sum_{i=1}^{k} u_{ij} = \sum_{i=1}^{k} \left(\frac{-\lambda_j}{m d^2(x_j, c_i)} \right)^{\frac{1}{m-1}} = 1 \tag{3.7}$$

$$\lambda_j = -\left(\sum_{i=1}^{k} \left(m d^2(x_j, c_i) \right)^{\frac{1}{1-m}} \right)^{1-m} \tag{3.8}$$

The new membership values are obtained by inserting λ_j into equation (3.6) yields:

$$u_{ij} = \left(\frac{\left(\sum_{l=1}^{k} \left(d^2(x_j, c_l) \right)^{\frac{1}{1-m}} \right)^{1-m}}{d^2(x_j, c_i)} \right)^{\frac{1}{m-1}} \tag{3.9}$$

$$u_{ij} = \frac{\left(d^2(x_j, c_i) \right)^{\frac{1}{1-m}}}{\sum_{l=1}^{k} \left(d^2(x_j, c_l) \right)^{\frac{1}{1-m}}} \tag{3.10}$$

Step 2: Optimizing the cluster centers:

The obtained membership values are used to optimize clusters centers by deriving of *J* with respect to centers. Thus, the cluster centers are updated by

$$c_i = \frac{\sum_{j=1}^{N} {u_{ij}}^m x_j}{\sum_{j=1}^{N} {u_{ij}}^m} \tag{3.11}$$

From a random initialization of clusters centers and using formulas (3.10) and (3.11), FCM algorithm recomputed clusters centers until no improvement of these centers. Once the clusters centers fixed, the algorithm assign each pixel x_i of the image to a cluster having maximum fuzzy membership degree.

Algorithm FCM

Input : K : number of cluster
N :number of pixel
ε: threshold
m: Fuzziness exponent (typically $m=2$).
Output : K clusters center C

1- Randomly select K initial cluster centers $C=\{c_1,\ldots,c_K\}$.
2- Update memberships U_{ij} using formula (10)
3- Calculate $J_{old}(u,c)$ using formula (1)
4- Update clusters center c_i using formula (11)
5- Calculate $J_{new}(u,c)$ using formula (1)
6- Repeat steps 2 to 5 until $|J_{new} - J_{old}| < \varepsilon$
7-Return C

2.1. Advantages of FCM

Fuzzy C-Means (FCM) is a widely used clustering algorithm that offers several key advantages over traditional hard clustering methods like K-Means. Below are its primary benefits, particularly in applications such as image segmentation:

1. Handles ambiguity and overlapping clusters: Unlike crisp clustering (which assigns each data point to only one cluster), FCM allows partial membership, meaning a point can belong to multiple clusters with varying degrees (between 0 and 1) which is useful in medical imaging (brain MRI segmentation where tissues overlap).
2. More flexible in noisy and uncertain data: Since FCM considers fuzzy membership values, it is less sensitive to minor noise compared to crisp clustering methods. Variants like Spatial FCM (SFCM) further improve noise robustness by incorporating neighborhood information.
3. Better for non-spherical and Complex data: Kernel FCM (KFCM, variant of FCM) can handle non-linear separability by mapping data into higher dimensions.
4. Adaptable with Customizable Fuzziness (m): The fuzzifier parameter (m) controls cluster overlap. If $m \rightarrow 1$, FCM behaves like K-Means (crisp clustering). But values more than 1 ($m > 1$) increases fuzziness (useful for uncertain data) which allows tuning based on application needs.
5. Works well in high-dimensional data: Effective in feature-rich datasets (hyperspectral images, gene expression data) and can be combined with dimensionality reduction (PCA) for efficiency.
6. Compatible with hybrid and bio-inspired optimizations: FCM can be enhanced with evolutionary algorithms (GA, PSO, ABC,…) to improve initial centroid selection, escape local optima and speed up convergence.

7. Wide range of applications: FCM clustering has diverse real-world applications due to its ability to handle uncertain data. In medical imaging, it aids tumor detection and tissue analysis, while in remote sensing; it enables precise land cover classification. The algorithm also proves valuable for industrial defect detection and computer vision tasks like object recognition, demonstrating its versatility across domains.

FCM's ability to model uncertainty, handle noise, and integrate with optimization techniques makes it a powerful tool for real -world clustering tasks- especially in image segmentation where data is often ambiguous. While it has higher computational costs than K-Means, its flexibility and accuracy justify its use in many applications.

2.2. Demerits of Fuzzy C-Means (FCM) Clustering

Despite its advantages, FCM has several limitations:

1. Sensitivity to Noise and Outliers: The standard FCM objective function weights all data points equally, making it vulnerable to corrupted or extreme values.

2. High Computational Cost: Iterative membership updates and distance calculations become expensive for large datasets.

3. Dependence on initial centroids: Poor initialization can lead to suboptimal clustering or slow convergence.

4. Assumes spherical clusters: Struggles with complex, non-linear, or irregular cluster shapes.

5. Requires predefined cluster number: Like K-Means, FCM needs the number of clusters (K) as input, which may be unknown in real-world data.

6. Parameter tuning (Fuzzifier m): Choosing an inappropriate m value can lead to overly fuzzy or rigid results.

3. FCM variants

The Fuzzy C-Means (FCM) algorithm has been extended into numerous variants to address its limitations, such as sensitivity to noise, outliers, and complex data structures. These adaptations include kernel-based FCM (KFCM), which maps data into higher-dimensional spaces for better separability, and weighted or entropy-regularized versions to improve robustness. Other variants incorporate spatial information, alternative distance metrics, or hybrid optimization techniques to enhance clustering accuracy and adaptability across diverse datasets. Bellow the most variants that marked the FCM evolutions.

3.1. Spatial FCM (SFCM)

Conventional FCM ignores spatial information, making it sensitive to noise and outliers in image processing or spatially correlated data. To address this, Spatial Fuzzy C-Means (SFCM) incorporates spatial constraints [Ahmed & Moriarty, 2002] [Chen et al., 2004], improving robustness in applications like medical image segmentation, remote sensing, and pattern recognition [Ali et al., 2023].

The objective function defined in formula (3.1) was modified [Chen et al., 2004] taking account of the spatial information in order to increase the robustness over noise as follow:

$$J(U,C) = \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_j, c_i) + \alpha \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(\bar{x}_j, c_i) \tag{3.12}$$

where $\bar{x}_j$ represents the grey value of pixel in the weighted averaging image window. α is a parameter to control the tradeoff between the original image and the corresponding mean-filtered image. Under the constraints defined in (3.2), the objective function in formula (3.12) can be optimized leading to a new algorithm called SFCM. Like the original FCM, SFCM, iteratively, computes clusters centers using the formulas below.

$$u_{ij} = \frac{(d^2(x_j, c_i) + \alpha d^2(\bar{x}_j, c_i))^{\frac{1}{1-m}}}{\sum_{l=1}^{K}(d^2(x_j, c_l) + \alpha d^2(\bar{x}_j, c_l))^{\frac{1}{1-m}}} \tag{3.13}$$

and

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^{m}(x_j + \alpha \bar{x}_j)}{(1+\alpha)\sum_{j=1}^{N} u_{ij}^{m}} \tag{3.14}$$

Advantages

The Spatial Fuzzy C-Means (SFCM) algorithm offers several key advantages over traditional FCM. First, its incorporation of spatial constraints enhances noise robustness, significantly reducing sensitivity to outliers and corrupted data points. Second, it preserves spatial continuity, making it particularly effective for tasks like image segmentation, where adjacent pixels often share cluster membership. Third, SFCM provides flexibility through the parameter *α*, which allows users to adjust the balance between spatial influence and feature-based clustering.

Limitations

Despite its strengths, SFCM has two primary limitations:

1- Higher computational cost due to the added neighborhood term, which increases algorithmic complexity,
2- Requires predefined cluster number: Like FCM, SFCM needs the number of clusters (*K*) as input, and may converge to local optima.

3- Parameter sensitivity, as performance heavily depends on the choice of α and the size of the spatial neighborhood.

3.2. Kernel-based FCM (KFCM)

The original FCM algorithm assumes that clusters are spherical and linearly separable in the input space. When this assumption fails the algorithm performs poorly. This limitation is caused by the use of the Euclidean norm metric. By using kernel methods, KFCM implicitly transforms the input data into a high-dimensional feature space where linear separation may be possible [Chang-Chien et al., 2021] [Abdullah, 2024].

Using kernel function, the KFCM objective function is expressed as:

$$J(U,C) = 2\sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m}(1 - Ker(x_j, c_i)) \tag{3.15}$$

where *Ker* is a kernel function. The Radial Basis Function (RBF) Kernel also called the Gaussian kernel is one of the most commonly used in KFCM.

Similarly to the FCM algorithm, this objective function can be optimized under the constraints defined in (2). We can compute the fuzzy membership function and the clusters centers with the formulas below respectively.

$$u_{ij} = \frac{\sum_{l=1}^{K}(1 - Ker(x_j, c_l))^{\frac{1}{m-1}}}{(1 - Ker(x_j, c_i))^{\frac{1}{m-1}}} \tag{3.16}$$

and

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} Ker(x_j, c_i) x_j}{\sum_{j=1}^{N} u_{ij}^{m} Ker(x_j, c_i)} \tag{3.17}$$

The general steps of KFCM are the same as FCM.

Advantages of KFCM

- Handles non-linear data separation: Uses kernel functions to map data into a higher-dimensional space, making it effective for complex, non-linear clusters and performs better than standard FCM when clusters are not well-separated in the original space.
- Robust to noise and outliers: Kernel methods can reduce the impact of noise by transforming data into a more separable space.
- Flexibility in kernel selection: Different kernels (RBF, polynomial, sigmoid) can be chosen based on the dataset, improving adaptability.
- Improved clustering accuracy: Often achieves better clustering results than FCM for datasets with intricate structures.
- Works well with high-dimensional Data: Kernel tricks help in dealing with the "curse of dimensionality" by implicitly working in a higher-dimensional feature space.

Limitations of KFCM

- Computationally expensive: Kernel matrix calculations require *O(N2)* memory and computations, making it slower than FCM for large datasets and not suitable for real-time or big data applications.
- Kernel parameter sensitivity: Performance heavily depends on kernel parameters (σ in the Gaussian kernel) which requires tuning via cross-validation or heuristic methods leading to time-consuming.
- Risk of overfitting : Poor kernel choices or parameter settings may lead to overfitting, especially with small datasets.
- Initialization Sensitivity : Like FCM, KFCM is sensitive to initial cluster centroids and may converge to local optima.
- Interpretability Issues: Since clustering occurs in a high-dimensional kernel space, interpreting results is harder than in linear methods like FCM.
- Not Always Better Than FCM for Simple Data: For linearly separable clusters, KFCM may introduce unnecessary complexity without significant gains.

3.3. Spatial Kernel-based FCM (SKFCM)

The Spatial Kernelized Fuzzy C-Means (SKFCM) is an extension of the Kernelized Fuzzy C-Means (KFCM) that incorporates spatial information from image or grid-based data to improve clustering performance, especially in noisy environments. Similar to SFCM, the spatial information is added to KFCM in the following way leading to a new algorithm [Raj, 2024]:

The objective function is formulated as (18):

$$J(U,C) = 2\sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m}(1 - Ker(x_j, c_i)) + 2\alpha \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m}(1 - Ker(\bar{x}_j, c_i)) \qquad (3.18)$$

where $\bar{x}_j$ represents the grey value of pixel in the weighted averaging image window.

Similar to standard FCM, the fuzzy membership matrix and the clusters centers are updated iteratively with the formulas (3.19) and (3.20) respectively.

$$u_{ij} = \frac{1}{\sum_{l=1}^{K}\left(\frac{(1-Ker(x_j,c_i))+\alpha(1-Ker(\bar{x}_j,c_i))}{(1-Ker(x_j,c_l))+\alpha(1-Ker(\bar{x}_j,c_l))}\right)^{\frac{1}{(m-1)}}} \qquad (3.19)$$

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^{m}(Ker(x_j,c_i)x_j + \alpha(1-Ker(\bar{x}_j,c_i))\bar{x}_j)}{\sum_{j=1}^{N} u_{ij}^{m}((1-Ker(x_j,c_i))+\alpha(1-Ker(\bar{x}_j,c_i)))} \qquad (3.20)$$

Advantages of SKFCM:

1. Handles non-linear data: Like KFCM, by using kernel functions, SKFCM can effectively cluster non-linearly separable data by mapping it to a higher-dimensional space.
2. Incorporates partial supervision: unlike traditional FCM, SKFCM can leverage labeled data (if available) to guide clustering, improving accuracy when some prior knowledge exists.
3. Robust to noise and outliers: The fuzzy membership approach allows soft clustering, making it less sensitive to noise compared to hard clustering methods like K-Means.
4. Flexible cluster shapes: The kernel trick enables the detection of arbitrarily shaped clusters, unlike standard FCM, which assumes spherical clusters.

Limitations of SKFCM:

1. Computational complexity: Kernel matrix computation is expensive *($O(n^2)$)*, making SKFCM slower than FCM for large datasets.
2. Parameter sensitivity: Performance depends on kernel selection (RBF, polynomial) and kernel parameters (σ in RBF), which require tuning.
3. Requires some labeled data: While semi-supervised, it still needs partial labels for optimal performance; fully unsupervised cases may not benefit as much.
4. Scalability Issues: Not suitable for big data applications due to high memory and computational demands.
5. Risk of overfitting: If the kernel parameters are poorly chosen, the model may overfit the training data.
6. Initialization sensitivity: Like FCM and KFCM, SKFCM is sensitive to initial cluster centroids and may converge to local optima.

3.4. Possibilistic Fuzzy C-Means (PFCM)

Possibilistic Fuzzy C-Means (PFCM) [Pal, 2005] [Farooq & Memon, 2024] is another extension of the standard Fuzzy C-Means (FCM) algorithm that fuzzy membership and possibilistic clustering (PCM) and addresses two key limitations:

1. Noise sensitivity: FCM forces all points to belong to clusters, making it vulnerable to outliers.
2. Membership interpretation: FCM's probabilistic constraints can lead to counterintuitive results.

PFCM introduces *possibilistic* memberships that represent the absolute degree of typicality of a point to a cluster and allow points to belong to no clusters (unlike FCM). The PFCM objective function combines two components: the fuzzy membership (u_{ij}: (similar to FCM)) and the possibilistic membership (t_{ij}: Measures typicality (like PCM))

$$J(U,C) = \sum_{i=1}^{K}\sum_{j=1}^{N}(au_{ij}^{m}+bt_{ij}^{\eta})d^2(x_j,c_i) + \sum_{i=1}^{K}\gamma_i\sum_{j=1}^{N}(1-t_{ij})^{\eta} \quad (3.21)$$

where:

K = number of clusters
N = number of data points
u_{ij} = fuzzy membership of x_j in cluster i (as in FCM)
t_{ij} = possibilistic typicality of x_j in cluster i (as in PCM)
c_i = centroid of the i^{th} cluster
a,b = weighting coefficients controlling the influence of fuzzy and possibilistic terms (a+b=1)
m = fuzzification exponent (m>1)
η = typicality exponent (usually η=2)
γ_i = scale parameter for the i^{th} cluster (similar to PCM)

The PFCM algorithm follows the steps bellow:

1. Choose K, m, η, a, b
2. Initialize cluster centers randomly
3. Repeat until convergence: centroids stabilize (change below a threshold) or max iterations reached.
    a. Update fuzzy memberships (u_{ij}) using formula (3.10)
    b. Update possibilistic memberships (t_{ij}) using formula (3.22)

$$t_{ij} = \left(1 + \left(\frac{b1 - d^2(x_j, c_i)}{\gamma_i}\right)^{\frac{1}{\eta-1}}\right)^{-1} \tag{3.22}$$

    c. Update cluster centers (c_i) using formula (3.23)

$$c_i = \frac{\sum_{j=1}^{N}(au_{ij}^m + bt_{ij}^\eta)x_j}{\sum_{j=1}^{N}(au_{ij}^m + bt_{ij}^\eta)} \tag{3.23}$$

Advantages of PFCM:

The PFCM algorithm offers several key advantages over traditional clustering methods.

1. Unlike FCM, PFCM is robust to noise and outliers, making it more reliable for real-world datasets with imperfections.
2. Additionally, it avoids coincident clusters, a common issue in PCM, by maintaining meaningful cluster separation.
3. PFCM effectively balances fuzzy membership and possibilistic typicality, allowing for better handling of uncertain data while preserving probabilistic interpretability.
4. This hybrid approach also makes it well-suited for overlapping clusters, where clear boundaries between groups are difficult to define.

Limitations of PFCM:

However, PFCM has some notable drawbacks.

1. It is highly sensitive to parameter choices, requiring careful tuning of coefficients (a, b, γ_i) to achieve optimal performance.
2. The algorithm also incurs a higher computational cost compared to FCM due to its combined membership and typicality calculations.
3. Furthermore, like many clustering methods, PFCM is initialization-dependent, meaning poor initial centroids can lead to suboptimal clustering results.

PFCM offers a robust alternative to FCM and PCM by combining their strengths, but its effectiveness depends heavily on proper parameter selection and initialization. It is well-suited for datasets with noise and overlapping clusters but requires careful tuning to achieve optimal results.

3.5. Improved FCM with Non-Local Information (FCM-NL)

To include non-local information, the Improved FCM-NL [Ma et al., 2014] [Zhang et al., 2017] [Zhang et al., 2021] modifies the objective function as follow:

$$J(U,C,W) = \sum_{i=1}^{K}\sum_{j=1}^{N}\left(u_{ij}^{m} d^2(x_j, c_i) + \lambda \sum_{l\in\rho_j} w_{jl}\, d^2(x_l, c_i) \right) \tag{3.24}$$

where:

U_{ij} is the membership that determines how much a pixel x_j belongs to cluster i

$$u_{ij} = \frac{1}{\sum_{p=1}^{K}\left(\frac{d^2(x_j,c_i) + \lambda\sum_{l\in\rho_j} w_{jl}\, d^2(x_l,c_i)}{d^2(x_j,c_p) + \lambda\sum_{l\in\rho_j} w_{jl}\, d^2(x_l,c_P)}\right)^{\frac{2}{m-1}}} \tag{3.25}$$

w_{jl} presents the non-local weight (measures similarity between patches (neighborhoods) around pixels j and l)

$$w_{il} = EXP\left(-\frac{d^2(P_j, P_l)}{\sum_{j=1}^{N} u_{ij}{}^{m}(1 + \lambda\sum_{l\in\rho_j} w_{jl}))} \right) \tag{3.26}$$

P_j and P_l are image patches centered at x_j and x_l.

c_i is the i^{th} cluster center, it is updated as a weighted average of all pixels, incorporating both intensity and non-local similarity:

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}{}^{m}\left(x_j + \lambda\sum_{l\in\rho_j} w_{jl}\, x_l\right)}{\sum_{j=1}^{N} u_{ij}{}^{m}(1 + \lambda\sum_{l\in\rho_j} w_{jl}))} \tag{3.27}$$

ρ_j is a search window around pixel j.

λ is a balancing parameter (controls influence of non-local term)

the FCM-NL algorithm is summarized as follow:

1. Initialize centroids c_i (randomly).
2. Repeat until convergence:
    - Compute non-local weights w_{jp} for all pixels.
    - Update membership values u_{ij}.
    - Update cluster centers v_i.

Advantages of FCM-NL:

The FCM with Non-Local Information (FCM-NL) algorithm significantly outperforms standard FCM in several key aspects.

1. Unlike traditional FCM, which relies solely on pixel intensity, FCM-NL incorporates patch-based similarity, making it highly robust to noise and outliers while preserving structural details.
2. This approach leverages non-local means filtering, effectively reducing blurring and maintaining sharp edges, a critical advantage in medical imaging (MRI and ultrasound) where fine details are essential.
3. FCM-NL achieves superior segmentation accuracy in noisy environments by adaptively smoothing homogeneous regions without degrading textures.
4. The integration of non-local information ensures adaptive noise suppression, making FCM-NL a powerful choice for real-world applications where noise corruption is inevitable.

Limitations of FCM-NL:

While FCM-NL improves noise robustness and segmentation accuracy over standard FCM, it has several key limitations:

1. High computational cost: Calculating non-local patch similarities (w_{jp}) is computationally expensive, especially for large images or 3D volumes and slower than standard FCM due to neighborhood search operations for every pixel.
2. Memory intensive: Storing patch-based weights for all pixel pairs requires significant RAM, limiting scalability.
3. Sensitive to parameter tuning: Performance depends heavily on Patch size (too small leads noise-sensitive; too large leads oversmoothing), smoothing parameter h (affects weight decay in w_{jl}) and trade-off parameter λ (balancing local vs non-local terms). Suboptimal choices of these parameters can lead to oversmoothing or inadequate noise removal.
4. Initialization sensitivity: Like FCM, results depend on initial cluster centers (poor initialization leads to suboptimal convergence).
5. The optimization landscape of FCM-NL (like FCM) is non-convex, meaning multiple local minima exist and the probabilistic memberships (summing to 1) impose constraints that may restrict movement toward a better global solution.

6. Limited adaptability to heterogeneous noise: Assumes uniform noise distribution; struggles with structured noise (salt-and-pepper, stripe artifacts).
7. Complex implementation: Requires additional steps (patch extraction, weight computation).

3.6. Weighted FCM (WFCM)

Weighted Fuzzy C-Means (WFCM) [Sarkar et al., 2024] [Poshitha et al., 2023] is an enhanced version of the standard FCM algorithm that incorporates feature weighting to improve clustering performance. Unlike traditional FCM, which treats all features equally, WFCM assigns different weights to features based on their importance, leading to more accurate and meaningful clustering.

The objective function of WFCM is defined as:

$$J(U,C,W) = \sum_{i=1}^{K}\sum_{j=1}^{N}\left(u_{ij}^{m}\sum_{l=1}^{D} w_l^{\beta}\, d^2(x_{lj}, c_{li})\right) \tag{3.28}$$

where

- $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$: Dataset with N samples.
- $\boldsymbol{x}_i=[x_{i1},x_{i2},\ldots,x_{id}]$: A sample with D features.
- K: Number of clusters.
- $C=\{c_1,c_2,\ldots,v_K\}$: Cluster centroids.
- $U=[u_{ij}]$: Fuzzy membership matrix.
- $W=[w_1,w_2,\ldots,w_D]$: Feature weights.
- β: Weight exponent (β>1, controls weight distribution).

WFCM minimizes the objective function $J(U, C, W)$ by optimizing iteratively U,V,W. It updates:

1. Fuzzy Memberships (u_{ij})

$$u_{ij} = \frac{1}{\sum_{i=1}^{K}\frac{\left(\sum_{l=1}^{D} w_l^{\beta}\, d^2(x_{lj}, c_{li})\right)^{\frac{1}{m-1}}}{\sum_{l=1}^{D} w_l^{\beta}\, d^2(x_{lj}, c_{lj})}} \tag{3.29}$$

2. Clusters centroids (c_{ij})

$$c_{il} = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_{lj}}{\sum_{j=1}^{N} u_{ij}^{m}} \tag{3.30}$$

3. Features weights (w_j)

$$w_l = \frac{1}{\sum_{l=1}^{D}\left(\frac{\sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_{lj},c_{il})}{\sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_{il},c_{il})}\right)^{\frac{1}{\beta-1}}} \quad (3.31)$$

The algorithm steps follow the general steps bellow:

1. Initialize:
   - Random cluster centroids *C*.
   - Uniform feature weights w_j.
2. Repeat until convergence:
   - Update memberships *U*.
   - Update centroids *C*.
   - Update feature weights *W*.
3. Terminate when:
   - Change in *U* or *J (U,C,W)* is below a threshold ϵ.

Advantages of WFCM:

Weighted Fuzzy C-Means (WFCM) offers several key advantages over traditional FCM, making it particularly effective for complex datasets:

- Feature selection: WFCM automatically detects and emphasizes important features by assigning higher weights to discriminative attributes while suppressing irrelevant ones. This leads to more meaningful clustering without manual feature engineering.
- Noise robustness: By reducing the influence of noisy or redundant features through adaptive weighting, WFCM improves robustness in real-world datasets where irrelevant variables may degrade performance.
- Better clustering accuracy: In high-dimensional data, WFCM outperforms standard FCM by focusing on the most relevant features, resulting in clearer cluster separation and higher accuracy.
- Flexibility: WFCM can be easily extended with kernel methods (for nonlinear data) making it adaptable to diverse applications.

Limitations of WFCM:

While WFCM improves upon traditional FCM, it has several key limitations:

- Sensitivity to initialization: Like FCM, WFCM's performance depends heavily on initial centroids and weights, leading to suboptimal solutions if poorly initialized.
- Computational Cost: The additional weight optimization step increases runtime, especially for high-dimensional data, making WFCM slower than standard FCM.
- Parameter Tuning Challenges: The weight exponent *(β)* and fuzziness parameter *(m)* require careful tuning. Poor choices can lead to overweighting/underweighting features or overly fuzzy clusters.

- Assumption of feature independence: WFCM treats features as independent, ignoring correlations. Real-world data with interdependent features may need kernel or graph-based extensions.
- Local optima trapping: The objective function is non-convex, so WFCM may converge to local optima, especially with noisy or overlapping clusters.
- Scalability issues: For very large datasets, WFCM's iterative weight updates become prohibitively expensive.

3.7. Entropy-Based FCM (EFCM)

Entropy-Based Fuzzy C-Means (EFCM) is a variant of FCM that incorporates entropy regularization to improve cluster validity and reduce sensitivity to initialization. Unlike traditional FCM, which relies solely on the sum of squared errors, EFCM introduces an entropy term to control the fuzziness of membership assignments, leading to more robust clustering [Kahali et al., 2019] [Ray & Sing, 2024].

EFCM introduces a regularization term that encourages more balanced and stable memberships. The modified objective function becomes:

$$J(U,C) = \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} d^2(x_j, c_i) - \lambda \sum_{i=1}^{K}\sum_{j=1}^{N} u_{ij}^{m} \log(u_{ij}) \tag{3.32}$$

- λ is the entropy regularization coefficient
- The entropy term $-\sum u_{ij} log(u_{ij})$ promotes high uncertainty or "softness" in the memberships, preventing premature hard clustering.

The optimization of this objective follows an iterative procedure:

1. Update Memberships:

$$u_{ij} = \frac{exp\left(-\frac{d^2(x_j, c_i)}{\lambda}\right)}{\sum_{l=1}^{K} exp\left(-\frac{d^2(x_j, c_l)}{\lambda}\right)} \tag{3.33}$$

2. Centroids update using formula (11) (as FCM)
3. Terminate when:
   Change in U or $J(U,C)$ is below a threshold ϵ.

Advantages of EFCM:

- Adaptive fuzziness: Automatically adjusts cluster fuzziness via entropy, reducing reliance on manual tuning of m.
- Robust to noise: Entropy regularization suppresses spurious membership assignments.
- Theoretical foundation: Derived from maximum entropy principle, ensuring mathematically sound membership distributions.

Limitations of EFCM:

While EFCM offers advantages over conventional FCM, it has several key limitations:

- Sensitivity to entropy coefficient (λ): The algorithm's performance heavily depends on proper selection of λ (λ too small: Behaves like hard K-means (loses fuzziness benefit), λ too large: Produces overly fuzzy clusters (near-uniform memberships)) and there is no universal rule for optimal λ selection.
- Computational complexity (Slower Convergence): Typically requires more iterations than standard FCM.
- Initialization sensitivity:
  - Centroid dependence: Like all FCM variants, results depend on initial centroids.
  - Local optima: May converge to suboptimal solutions.
- Cluster shape assumptions: Inherited FCM Limitations (Still assumes hyper-spherical clusters).

4. Discussion

Fuzzy C-Means (FCM) clustering has evolved into numerous variants to address its inherent limitations, such as sensitivity to noise, dependence on initial centroids, and difficulty handling complex data structures. Key variants include Possibilistic FCM (PFCM) and Non-Local FCM (NL-FCM) AND EFCM for robustness against outliers, Kernel FCM (KFCM) for nonlinear data separation, and Spatial FCM (SFCM) for image processing tasks. Despite these improvements, standard FCM still faces challenges like local optima convergence, sensitivity to initialization and need of clusters number. This is where bio-inspired optimization methods like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC) prove invaluable. These methods enhance FCM by automating centroid initialization, dynamically optimizing parameters, and escaping local optima through global search strategies. Such hybrid approaches are particularly effective in medical imaging, pattern recognition, and big data clustering, where traditional FCM struggles. The integration of bio-inspired optimization with FCM not only improves clustering accuracy but also reduces computational costs, making it a powerful tool for complex real-world applications.

5. Optimization methods

Optimization is a discipline focused on identifying the most effective solution from a range of possible options to address a specific problem, leveraging mathematical and computational methods. At its core, it entails optimizing (maximizing or minimizing) an objective function subject to certain constraints [Nesterov, 2018].

Optimization problems are ubiquitous in many fields, such as engineering, logistics, finance, physics, chemistry, and more.

5.1. Categories of optimization

Optimization techniques can be broadly categorized into deterministic and stochastic approaches.

a) **Deterministic Optimization** is a class of optimization methods that operates under the assumption that all input data is precisely known and free from randomness. It relies on exact mathematical models, such as linear or nonlinear equations, to find optimal solutions . It includes:
   1. Linear Programming (LP): Optimizes a linear objective function subject to linear constraints, widely used in resource allocation and scheduling.
   2. Nonlinear Programming (NLP): Deals with nonlinear relationships in the objective or constraints, common in engineering design and economics.
   3. Integer Programming: Restricts decision variables to discrete (integer) values, essential for problems like logistics and network design.
   4. Convex Optimization: Focuses on convex objective functions and constraint sets, enabling efficient global solutions in machine learning and control systems.

These methods are favored when system parameters are well-defined and uncertainty is negligible.

b) Stochastic optimization: known as Non-deterministic methods, refers to methods that handle problems where uncertainty, randomness, or incomplete information plays a significant role. Unlike deterministic approaches (which assume exact, fixed inputs), these techniques incorporate probabilistic models, heuristics, or adaptive strategies to find robust or approximate solutions [Spall, 2005].

5.2. Heuristic optimization methods

They are intelligent approaches designed to find good approximate solutions in a reasonable time, especially for complex, nonlinear, or NP-hard problems. Unlike exact methods, they do not guarantee optimality but are flexible and adaptable to various real-world challenges.

5.2.1. Bio-Inspired methods

Bio-inspired methods, Known also as Nature-inspired metaheuristics, are optimization algorithms modeled after biological, ecological, or social behaviors observed in nature. These methods excel in solving complex, non-linear, and high-dimensional problems where traditional approaches struggle. They are broadly classified into three categories:

- Evolutionary Algorithms (EA): These mimic biological evolution through mechanisms like selection, recombination, and mutation. in this category we distinguish Genetic Algorithm (GA) [Holland, 1992], and Evolution Strategies (ES) [Storn & Price, 1997]. Inspired by Darwinian natural selection, GA uses *crossover*, *mutation*, and *fitness-based selection* to evolve solutions, where ES, A variant of GA, emphasizes *self-adaptation* of mutation parameters. It is particularly effective in continuous optimization.

- Swarm Intelligence (SI): These algorithms simulate collective behavior in decentralized systems. In this category, we cite Particle Swarm Optimization (PSO) [Kennedy & Eberhat, 1995] that models bird flocking or fish schooling, where "particles" adjust their trajectories based on individual and group best positions and Ant Colony Optimization (ACO) [Dorigo et al., 1996], method based on pheromone trail deposition by ants.
- Collective Intelligence Algorithms: These leverage group behavior for exploration-exploitation trade-offs. We distinguish Artificial Bee Colony (ABC) [Karaboga & Basturk, 2007] that simulates honeybee foraging with "employed," "onlooker," and "scout" bees balancing exploration and exploitation and Firefly Algorithm (FA) [Yang, 2009] which is inspired by firefly bioluminescence. FA uses attractiveness-based movement for multi-modal and multi-objective problems.

5.2.2. Physics/Chemistry-Based Methods

Physics and chemistry-inspired optimization methods leverage natural phenomena such as thermal dynamics, gravitational forces, and molecular interactions to solve complex optimization problems. These algorithms mimic processes like annealing in metallurgy, or chemical reactions to explore solution spaces efficiently. By translating physical laws into computational strategies, they offer robust alternatives to traditional mathematical optimization, particularly in high-dimensional, nonlinear, or multimodal problems. In this category, we cite:

- Simulated Annealing (SA): inspired by thermodynamics (gradual cooling). Its principle is to accept worse solutions temporarily to escape local optima [Kirkpatrick et al., 1983] [Guilmeau et al., 2021].
- Harmony Search (HS): mimics musical improvisation where musicians adjust their pitches to achieve a pleasing harmony. It is proposed by Geem, Kim, and Loganathan in 2001 [Geem et al., 2001].
- Gravitational Search Algorithm (GSA): based on the law of gravity and mass interactions, proposed by Rashedi, Nezamabadi-pour, and Saryazdi in 2009 . It mimics Newtonian physics, where masses (solutions) attract each other due to gravitational force, leading to global optimization [Mittal et al., 2021].

5.2.3. Local and Guided Search Methods

Local search methods focus on improving a solution by exploring its immediate neighborhood. They are efficient but may get stuck in local optima, while Guided Search methods combine global exploration with local refinement.

- Hill Climbing (Greedy Local Search): its concept is to iteratively move to the best neighboring solution.
- Tabu Search (TS): is an advanced local search algorithm that enhances hill climbing by using adaptive memory to escape local optima. It is introduced

by Fred Glover in 1986. It uses memory (tabu list) to avoid revisiting solutions [Prajapati et al., 2020].

- Variable Neighborhood Search (VNS): it systematically explores different neighborhood structures to escape local optima. It is proposed by Hansen and Mladenović in 1997 [Hansen & Mladenović, 2001].
- Iterated Local Search (ILS): combines local search with periodic perturbations to escape local optima. It operates by repeatedly:
    1. Applying local search to reach a local optimum
    2. Perturbing the current solution to escape the local optimum
    3. Repeating the process to explore the search space effectively
- Guided Local Search (GLS): is an intelligent local search metaheuristic that enhances traditional local search by dynamically modifying the objective function to escape local optima. It is developed by Voudouris and Tsang in the 1990s [Voudouris & al., 2010].

5.2.4. Hyper-Heuristics Methods

Hyper-heuristics are high-level search methodologies that automate the selection, combination, or generation of simpler heuristics (or components of heuristics) to solve complex optimization problems. They operate on a "heuristic space" rather than directly on the solution space, making them highly flexible and adaptable across different problem domains [Dokeroglu & al., 2024].

Hyper-heuristics can be classified into two main categories:

1. Selection Hyper-Heuristics that choose or switch between existing heuristics during the search process like Simple Random (randomly selects heuristics), Greedy Selection (picks the best-performing heuristic) or Markov Chain-based (uses transition probabilities).
2. Generation Hyper-Heuristics that generate new heuristics or heuristic components automatically such as Genetic Programming-based (evolves heuristic rules) and Neural Program Synthesis (deep learning-based heuristic generation).

6. FCM Optimization based on Bio-Inspired Methods

As mentioned above, FCM clustering has proven valuable for pattern recognition, medical imaging and data analysis, but its effectiveness is limited by four key challenges: (1) high sensitivity to initial centroid selection, (2) number of clusters, (3) tendency to converge to local optima, and (4) degraded performance with noisy or high-dimensional datasets. To address these limitations, researchers have successfully integrated bio-inspired optimization algorithms with FCM, yielding significant improvements in clustering performance. These nature-inspired computational techniques, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and others methods, enhance FCM through three primary mechanisms: (i) intelligent initialization of cluster centroids, (ii) dynamic refinement of membership functions, and (iii) robust global search capabilities that prevent premature convergence. The evolution of these hybrid approaches has followed a clear trajectory in

computational intelligence research, with each successive generation of bio-inspired methods offering enhanced capabilities for FCM optimization. In the following section, we examine key developments in this field, highlighting how various biological optimization paradigms have advanced the state-of-the-art in fuzzy clustering.

6.1. Genetic Algorithm

The Genetic Algorithm (GA) is a metaheuristic optimization algorithm inspired by the process of natural evolution [Holland, 1975]. It is based on the principle that solutions best adapted to a given problem are more likely to survive and reproduce, passing their characteristics to subsequent generations. A comprehensive can be found in [Katoch, 2021].

The operation of a genetic algorithm can be summarized as follows:

- Initialization: A starting population of solutions is randomly generated.
- Evaluation: The objective function of the problem is computed for each solution.
- Selection: The fittest solutions are selected for reproduction.
- Crossover: Selected solutions are combined to produce new offspring solutions.
- Mutation: New solutions may undergo random mutations to introduce diversity.
- Replacement: The new solutions replace the least fit individuals in the population.

These steps are repeated iteratively until either an optimal solution is found or a predetermined number of generations is reached.

Genetic Algorithms (GA) significantly enhance Fuzzy C-Means (FCM) clustering by addressing its key limitations, such as sensitivity to initial cluster centroids and tendency to converge to local optima. GA improves FCM through global search capabilities, where genetic operators like selection, crossover, and mutation explore the solution space more effectively than traditional random initialization. By optimizing cluster centers and membership matrices, GA-FCM hybrids achieve better convergence accuracy. Additionally, GA can automatically determine the optimal number of clusters by optimizing validity indices, eliminating the need for manual selection.

The authors in [Maulik & Bandyopadhyay, 2003] introduced an algorithm called Fuzzy Partitioning Using a Real-Coded Variable-Length Genetic Algorithm (FVGA) to automatically determine the optimal number of clusters along with their fuzzy clustering results. In FVGA, they employed a genetic algorithm (GA) combined with the Xie-Beni cluster validity index as a fitness function to guide chromosome evolution.

Building on this work, Saha and Bandyopadhyay [Saha & Bandyopadhyay, 2009] proposed a fuzzy dynamic clustering algorithm called the Fuzzy Variable-Length Genetic Algorithm with Point Symmetry (Fuzzy-VGAPS). In their approach, they incorporated a point symmetry-based validity measure, termed the fuzzy Sym-index, as the objective function for clustering.

In [Jansi & Subashini, 2014] and [Das & De, 2017], GA is used to optimize the initial clustering center firstly, and then FCM and KFCM algorithm respectively are availed to guide

the categorization, so as to improve the clustering performance of the FCM and KFCM algorithms.

Dong et al. [Dong et al., 2018] developed an adaptive fuzzy clustering approach that integrates Fuzzy C-Means (FCM) with a multi-objective genetic algorithm. Their method eliminates the need for predefined cluster numbers by employing an evolutionary optimization framework. This adaptive mechanism automatically determines the right number of clusters.

6.2. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a population-based meta-heuristic algorithm that mimics the collective intelligence observed in natural swarms, such as bird flocks or fish schools [Kennedy & Eberhart, 1995].

The Key Components of this algorithm are:

- Particles: Potential solutions that move through the search space
- Swarm: The collection of all particles
- Velocity: Determines particle movement direction and speed
- *pBest*: A particle's personal best solution found
- *gBest*: The swarm's global best solution found

Based on these components, the algorithm performs the following steps:

1. Initialization: A population of candidate solutions (particles) is randomly generated within the problem's search space.
2. Particle Movement: Each particle moves through the search space based on two key factors:
    - Its own best-known position (*pbest*)
    - The swarm's global best-known position (*gbest*)
3. Position Update: The velocity and position of each particle are adjusted using these two values, steering the swarm toward potentially better solutions.
4. Fitness Evaluation: The objective function evaluates each particle's quality (fitness).
5. *gbest* Update: If a particle discovers a solution superior to the current *gbest*, the *gbest* is updated.
6. Iteration: This process repeats for a set number of iterations, allowing the swarm to converge toward optimal or near-optimal solutions.

Particle Swarm Optimization (PSO) offers several key advantages that contribute to its popularity in optimization problems. First, it is straightforward to implement, requiring only a few parameters while maintaining a simple conceptual framework that can be easily adapted to various problem types. Second, PSO demonstrates strong global search capabilities through its unique combination of personal best (*pbest*) and global best (*gbest*) mechanisms, enabling effective exploration of the search space while avoiding local optima. Additionally, the algorithm exhibits robustness against environmental variations, showing minimal sensitivity to initial parameter settings and changing conditions. Finally, PSO boasts remarkable

versatility, as it can be successfully applied to a wide range of optimization challenges, including continuous, discrete, and multi-objective problems, making it a valuable tool across numerous scientific and engineering domains.

PSO is successfully applied to overcome the shortcomings of FCM. In [Liu & al., 2008] and [Izakian & Abraham, 2011], authors used PSO to overcome the problem of local minima. Kang and Zhang [Kang, 2012] proposed a hybrid clustering approach that combines FCM with PSO clustering problem. Their PSO-FCM algorithm addresses two key limitations: (1) it prevents FCM from converging to local optima through PSO's global search capabilities, while (2) simultaneously overcoming PSO's characteristic slow convergence by leveraging FCM's efficient local search. This synergistic integration demonstrates improved performance in clustering tasks compared to using either method independently. Also in [SK, 2021], PSO is successfully applied with FCM for leaf diseases prediction and in [Pham, 2018] [Verma, 2021] to segment brain image overcoming the local optima FCM's limit.

In [Tan et al., 2023], authors tackle the FCM initialization problem for image segmentation task leading to optimal initialization, thus faster segmentation.

6.3. Ant Colony Optimization

Ant Colony Optimization (ACO) is a metaheuristic optimization technique inspired by the foraging behavior of ants [Dorigo et al., 1996]. It is particularly effective for solving combinatorial optimization problems such as the traveling salesman problem, knapsack problem, and machine scheduling problems.

The fundamental principle of ACO relies on pheromone trail communication. Ants deposit pheromones along their paths, which serve as indicators of path quality for subsequent ants.

In ACO implementation:

1. A population of artificial ants is randomly initialized.
2. Each ant constructs a solution (path) by probabilistically following pheromone trails.
3. Solution quality is evaluated based on objective criteria (path length or value).

after each iteration:

- Pheromone trails are updated, with stronger reinforcement given to higher-quality solutions
- This positive feedback mechanism progressively biases the colony toward optimal paths

The algorithm iterates until convergence to an optimal or near-optimal solution is achieved.

Several studies have explored the hybridization of ACO and FCM to enhance clustering performance by optimizing FCM parameters, particularly:

- Cluster centers initialization (avoiding local optima)
- Optimal cluster number (eliminating the need for predefined number)

Shelokar et al. [Shelokar et al., 2004] introduced an ACO-FCM hybrid approach where ACO is employed to optimize the initial cluster centers before applying FCM for final refinement. While Saha and Sanghamitra [Saha, 2010] used ACO to determine the right number of cluster in unsupervised fuzzy data clustering task ensuring cluster compactness separation between clusters.

In [Wang & al, 2012] and [Raghtate & Salankar, 2015], authors used ACO to tackle the problem of local optima in fuzzy image segmentation task and in [Kumar & al., 2024] to optimize routing in flying Ad-Hoc network.

6.4. Bat Algorithm

The Bat Algorithm (BA) is a bio-inspired metaheuristic optimization technique developed by Xin-She Yang [Yang, 2010]. It mimics the echolocation behavior of bats, which use ultrasonic pulses to detect prey, avoid obstacles, and navigate in darkness. The algorithm efficiently balances exploration (global search) and exploitation (local refinement) by adjusting frequency, loudness, and pulse emission rates.

The BA Works as follows:

1. Initialization:
   - A population of bats (potential solutions) is randomly generated within the search space.
   - Each bat is assigned a position, velocity, frequency, and loudness.
2. Movement:
   - Bats move through the search space based on their velocity and frequency.
   - Frequency is adjusted using a frequency-tuning technique to explore different regions.
   - Loudness gradually decreases to focus on more promising areas.
3. Local Search:
   - Each bat generates a new solution randomly in a local search area around its current position.
   - If the new solution is better, it replaces the previous one.
4. Evaluation:
   - The quality (fitness) of each bat's solution is evaluated.
5. Update:
   - The velocities and positions of bats are updated based on their frequencies, loudness, and the best solutions found so far.
   - Loudness decreases as bats approach potential prey (optimal solutions).
6. Iteration:
   - Steps 2 to 5 are repeated for a set number of iterations or until a satisfactory solution is found.

The Bat Algorithm (BA) stands out as a highly efficient metaheuristic optimization technique due to several distinctive advantages. One of its most notable strengths is its balanced exploration and exploitation mechanism, achieved through dynamically adjustable frequency and loudness parameters. The frequency governs the search range, while loudness and pulse emission rate systematically shift focus from global exploration to local refinement as the algorithm progresses.

The Bat Algorithm (BA) significantly improves the performance of Fuzzy C-Means (FCM) clustering through two key mechanisms. First, it enables automatic cluster center initialization by optimizing the initial centroids, thereby reducing FCM's sensitivity to random initialization and improving solution consistency [Boulanouar & Lamiche, 2020] [Alhassan & Wan Zainon, 2020]. Second, BA enhances convergence and accuracy through its global search capabilities, which help FCM avoid local optima traps and produce more reliable clustering results [Jai, 2021]. These combined improvements make BA-FCM hybrids particularly effective for complex clustering tasks where traditional FCM struggles with initialization dependency and suboptimal convergence.

6.5. Artificial Bee Colony (ABC) Algorithm

The Artificial Bee Colony (ABC) algorithm is a swarm intelligence optimization technique inspired by the foraging behavior of honeybees [Karaboga & Basturk, 2007]. It is designed to solve complex optimization problems across various domains, particularly those involving high-dimensional variables and nonlinear objective functions.

The ABC Algorithm operates as follows

1. Initialization:
   A population of potential solutions ("food sources") is randomly generated.
2. Employed Bees:
   Each solution is assigned an "employed bee." These bees exploit their designated food source by searching its neighborhood for improved solutions.
3. Onlooker Bees:
   Onlooker bees select food sources based on fitness values and a "waggle dance" communication (where employed bees share discovery information). They then modify these sources through exploration, potentially identifying better solutions.
4. Scout Bees:
   If a food source shows no improvement after a predefined number of iterations, it is abandoned. A scout bee then randomly searches for a new food source.
5. Selection and Replacement**:**
   The best solutions discovered by employed and onlooker bees are retained, replacing abandoned sources.
6. Iteration:
   The process repeats for a fixed number of iterations or until a satisfactory solution is found.

The Artificial Bee Colony (ABC) algorithm significantly improves Fuzzy C-Means (FCM) clustering by addressing two critical limitations: sensitivity to initial cluster centroids and tendency to converge to local optima. ABC's unique three-phase search mechanism - employed bees for local exploitation, onlooker bees for solution refinement, and scout bees for global exploration - provides a robust framework for optimizing FCM's initial cluster centers [Karaboga & Ozturk, 2011]. The employed bees' neighborhood search helps fine-tune centroid positions, while scout bees prevent stagnation by randomly exploring new solutions

when improvements plateau [Alrosan & Norwawi, 2017]. This hybrid approach (ABC-FCM) demonstrates superior performance in cluster validity indices (Xie-Beni, Davies-Bouldin) compared to standard FCM, particularly for high-dimensional datasets where traditional FCM often fails [Lingappa & al., 2018]. Furthermore, ABC's ability to maintain population diversity through its abandonment-replacement mechanism enables more comprehensive search space exploration, resulting in more accurate and stable clustering solutions [Alomoush, 2022a] and his ability to be hybridized with others technics [Ni, 2024].

6.6. Firefly Algorithm

The Firefly Algorithm (FA), introduced by Yang [Yang, 2009], is a bio-inspired metaheuristic optimization technique that mimics the flashing behavior and social interactions of fireflies. This algorithm is particularly effective for solving complex multimodal optimization problems by simulating how fireflies are attracted to brighter light sources, which represent better solutions in the search space.

FA operates based on three key idealized rules:

1. Attraction Principle: All fireflies are unisex, and less bright fireflies move toward brighter ones.
2. Brightness-Distance Relationship: The attractiveness between fireflies decreases with increasing distance.
3. Objective-Dependent Brightness: A firefly's brightness is determined by the landscape of the objective function.

The FA can be summed up as follows.

1. Initialization:
   - Generate initial population of fireflies
   - Evaluate initial brightness (objective function)
2. Main Loop:
   - For each firefly, compare with all others
   - Move less bright fireflies toward brighter ones
   - Update positions with attractiveness and randomization
   - Re-evaluate brightness
3. Termination:

   Repeat until stopping criteria met (max iterations or convergence).

FA improves FCM by optimizing the initial cluster centroids, overcoming FCM's sensitivity to random initialization. The algorithm's attraction mechanism helps identify promising regions in the search space, leading to better starting points for FCM iterations [Kumar & Kumari, 2018].

FA-FCM hybrids can dynamically determine the optimal number of clusters by leveraging Firefly Algorithm’s (FA) global search capabilities alongside cluster validity indices. Unlike traditional FCM, which requires manual selection of the number of cluster, FA optimizes validity indices to identify the best cluster count. This eliminates subjectivity in cluster number-selection, enhancing automation and robustness in clustering tasks.

The integration of FA with FCM enhances clustering performance by leveraging FA's global search capability, which helps FCM escape local optima which is the common limitation of standard FCM. FA's attraction-repulsion mechanism dynamically balances exploration (searching new regions) and exploitation (refining existing solutions), ensuring a more efficient convergence toward optimal cluster centroids. Empirical studies demonstrate that FA-FCM hybrids achieve faster convergence and higher accuracy compared to traditional FCM, particularly in complex or high-dimensional datasets where FCM alone stagnates in suboptimal solutions [Alomoush, 2022b] [Thomas & Kumar, 2024].

6.7. Gray Wolf Optimizer

The Gray Wolf Optimizer (GWO) is a metaheuristic optimization technique inspired by the social hierarchy and hunting behavior of gray wolves [Mirjalili et al., 2014]. It simulates the leadership and cooperative hunting strategies of wolf packs to efficiently explore the search space and find optimal solutions to complex problems.

The Gray Wolf Optimizer operates as follows:

1. Initialization:
   - A population of wolves (potential solutions) is randomly initialized within the search space.
   - Four wolves are designated as the alpha (best solution), beta (second-best), delta (third-best), and omega (remaining wolves).
2. Search:
   - The alpha, beta, and delta wolves guide the search:
     - The alpha moves randomly to explore new promising regions.
     - The beta and delta refine solutions by gradually approaching the alpha's position.
   - The omega wolves follow the leaders and adjust their positions based on the hierarchy.
3. Attack (Exploitation):
   - If a better solution than a current leader is found, it replaces that leader in the hierarchy (solution).
4. Update:
   - The wolves' positions are updated based on their roles and the leaders' positions.
   - The hierarchy is dynamically adjusted according to solution quality.
5. Iteration:
   - Steps 2–4 are repeated for a predefined number of iterations or until a satisfactory solution is found.

The hierarchy-driven search mechanism in GWO ensures an effective balance between exploration (led by the alpha wolf's global search) and exploitation (guided by beta and delta wolves' local refinement), preventing premature convergence. This structure is enhanced by adaptive leadership, where the hierarchy dynamically updates when better

solutions emerge, ensuring continuous improvement. Additionally, the omega wolves' subordinate role maintains population diversity, promoting efficient convergence across complex search spaces.

The GWO significantly improves the performance of FCM clustering by addressing its two major limitations: sensitivity to initial centroids and tendency to converge to local optima. By optimizing the initial cluster centroids before FCM refinement, GWO ensures a more robust initialization, reducing dependency on random starting points [Katarya & Verma, 2018]. Furthermore, GWO's adaptive leadership update dynamically refines cluster centers during iterations, enhancing convergence accuracy. Empirical studies show that the hybrid GWO-FCM achieves superior results compared to standard FCM, particularly in complex datasets where traditional FCM fails to identify optimal partitions [Mohammdian-Khoshnoud et al., 2022]. This synergy combines FCM's local search precision with GWO's global optimization strength, yielding faster convergence.

7. Summary

The integration of Fuzzy C-Means (FCM) with bio-inspired optimization algorithms has emerged as a powerful approach to overcome FCM's limitations of sensitivity to initialization and local optima convergence. Metaheuristics like Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Artificial Bee Colony (ABC), and Firefly Algorithm (FA) enhance FCM by optimizing initial cluster centroids through global search mechanisms inspired by natural behaviors. These hybrid systems combine FCM's local search precision with bio-inspired algorithms' exploration capabilities, typically improving clustering accuracy while maintaining interpretability. The hybridization framework generally follows a two-phase process: bio-inspired methods first identify promising centroid positions, which FCM then refines through iterative minimization of the objective function. This synergistic approach has proven particularly effective in complex domains like medical image segmentation and high-dimensional data clustering, where conventional FCM often underperforms.

Key benefits include:

- Robustness to initialization
- Escape from local optima
- Automatic cluster number determination
- Improved convergence rates

Particularly, The Artificial Bee Colony (ABC) algorithm, inspired by the foraging behavior of honeybees, can significantly enhance the performance of Fuzzy C-Means (FCM) clustering. While FCM is effective for soft clustering, it suffers from sensitivity to initial centroids and a tendency to converge to local optima. ABC helps mitigate these issues by optimizing the initial cluster centers before FCM refinement. The ABC algorithm employs three types of bees -employed, onlooker, and scout bees- to balance exploration and

exploitation. Employed bees search for solutions (centroids), onlooker bees probabilistically select promising solutions, and scout bees introduce randomness to avoid stagnation. In the hybrid ABC-FCM approach, ABC first searches for optimal initial centroids by minimizing FCM's objective function, then FCM fine-tunes the membership degrees and cluster assignments. The table below presents a detailed comparison between ABC and other bio-inspired optimization techniques.

Table 3.1: ABC vs other bio-inspired methods

Category	Artificial Bee Colony (ABC)	Other Methods
Inspiration	Honeybee foraging (employed, onlooker, scout bees)	PSO: Bird flocking ACO: Ant pheromones GA: Natural selection FA: Firefly flashes GO: Hierarchy and hunting behavior of grey wolves (α, β, δ)
Search Mechanism	Three phases: employed, onlooker, scout bees	PSO: Velocity updates ACO: Probabilistic path selection GA: Crossover/mutation FA: Attraction-based movement GO: Three phases: Encircling, hunting, attacking prey - Guided by alpha (best), beta, and delta wolves
Exploration	High (scout bees enable random jumps)	PSO: Moderate ACO: High (pheromone evaporation) GA: High (mutation) FA: High (automatic subdivision) GO: Moderate (hierarchical leadership guides search)
Exploitation	Moderate (onlooker bees refine solutions)	PSO: High (fast convergence) ACO: High (positive feedback) GA: Moderate FA: Moderate GO: Very High (precise attacking phase near prey)
Key Parameters	Colony size, abandonment limit	PSO: Inertia weight (w), \(c_1, \(c_2 ACO: alpha, beta, evaporation rate GA: Crossover/mutation rates FA: beta_0, gamma GO: Convergence parameter (`a`) - Population size
Strengths	Balances exploration/exploitation; robust	PSO: Simple/fast ACO: Best for discrete problems GA: Flexible FA: Multi-modal optimization GO: Few parameters, easy to implement - High precision in local search - Good for unimodal problems
Weaknesses	Slow convergence in high dimensions	PSO: Premature convergence ACO: Parameter-sensitive GA: Computationally heavy FA: Distance metric reliance GO: May over-exploit local optima - Less effective in multimodal problems
Best For	Continuous optimization (engineering design) Multimodal optimization, Neural network training, Complex and noisy search spaces	PSO: Continuous spaces ACO: Combinatorial GA: Complex search spaces FA: Image processing GO: Unimodal optimization - Parameter tuning - Engineering design (where precision matters)

This hybridization offers key advantages, including better avoidance of local optima and robustness to initialization, making it particularly useful for high-precision tasks like medical image segmentation. However, ABC-FCM is computationally slower than some alternatives, such as PSO-FCM or GWO-FCM, due to ABC's inherent complexity. For improved efficiency, ABC can be combined with faster optimizers like PSO in a two-phase hybrid model -using PSO for a quick initial search and ABC for refinement. Despite its slower convergence, ABC-FCM remains a strong choice when clustering accuracy is prioritized over speed, especially for small to medium-sized datasets which the case of medical image segmentation.

8. Conclusion

This chapter has explored the hybridization of Fuzzy C-Means (FCM) clustering with bio-inspired optimization methods to overcome its inherent limitations of sensitivity to initialization and susceptibility to local optima. Various bio-inspired algorithms, including Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Firefly Algorithm (FA), and Artificial Bee Colony (ABC) and other methods have been examined for their ability to enhance FCM by optimizing initial cluster centroids and guiding the search toward globally optimal solutions. While each method offers distinct advantages, the Artificial Bee Colony (ABC) algorithm emerges as particularly superior for hybridizing with FCM due to its exceptional balance of exploration and exploitation, adaptability to high-dimensional spaces, and efficient convergence properties.

ABC's unique mechanisms, employed bees for local refinement, onlooker bees for solution selection, and scout bees for escaping local optima, make it more robust and reliable than PSO, GA, or FA when combined with FCM. Empirical studies consistently demonstrate that ABC-FCM achieves higher clustering accuracy, faster convergence, and better stability across diverse datasets, including noisy and complex real-world applications such as medical imaging and pattern recognition. While other bio-inspired methods also improve FCM, ABC's self-adaptive search strategy and reduced parameter dependency make it the most effective choice for enhancing FCM's performance.

ABC-FCM is favored Over other Hybrids for:

✓ Better Exploration-Exploitation Balance
✓ Fewer Control Parameters
✓ Superior Local Optima Avoidance
✓ Proven Higher Accuracy

Chapter 4 : Hybrid FCM-ABC Method for Medical Image Segmentation

1. Introduction

Image segmentation plays a pivotal role in medical imaging, particularly in brain MRI analysis, where accurate delineation of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is essential for clinical diagnosis and research. Among the various segmentation techniques, Fuzzy C-Means (FCM) clustering has been widely adopted due to its ability to handle the inherent ambiguity in tissue boundaries. However, traditional FCM suffers from several critical limitations: (1) sensitivity to initial cluster centers, (2) dependence on a predefined number of clusters, and (3) susceptibility to local minima convergence, especially in the presence of noise and intensity inhomogeneities. These shortcomings often lead to suboptimal segmentation, necessitating the development of more robust approaches.

To address these challenges, this chapter introduces a novel hybrid method that combines FCM with the Artificial Bee Colony (ABC) optimization algorithm. The ABC algorithm, inspired by the foraging behavior of honeybee colonies, is a powerful metaheuristic known for its global search capabilities, adaptability, and robustness in solving complex optimization problems. By integrating ABC with FCM, the proposed Hybrid FCM-ABC method mitigates the weaknesses of conventional FCM. Specifically, ABC optimizes the initial cluster centers, dynamically adjusts the number of clusters, and avoids local minima through its explorative search mechanism. This hybridization not only enhances segmentation accuracy but also improves computational efficiency and noise resilience.

The chapter begins by presenting the biological foundations of bee colony behavior and its artificial counterpart, the ABC algorithm, highlighting its key components -employed bees, onlooker bees, and scout bees- and their roles in optimization. Next, the hybrid FCM-ABC framework is detailed, explaining how ABC's global search capabilities are leveraged to refine FCM's clustering process. The experimental validation is conducted on both simulated brain MRI images (with controlled noise and intensity variations) and real clinical MRI datasets. Comparative analyses against standard FCM, Genetic Algorithm-based FCM (GA-FCM), and FCM with Covariance Matrix Adaptation Evolution Strategy (FCMA-ES) and other methods demonstrate the superiority of the proposed method in terms of segmentation accuracy, robustness to noise, and computational stability.

The results underscore the clinical relevance of Hybrid FCM-ABC method, particularly in scenarios where noise and artifacts compromise traditional methods. By overcoming the pitfalls of FCM while maintaining its interpretability, the proposed approach offers a promising tool for automated brain tissue segmentation, with potential applications in neurodegenerative disease diagnosis, surgical planning, and longitudinal studies. This chapter lays the groundwork for future research directions, including the integration of deep learning for further refinement and extension to other medical imaging modalities.

2. Biological Bee Colony

A honeybee colony is a highly organized social system where individual bees perform specialized roles to ensure the survival and efficiency of the hive. The colony consists of three primary castes, each with distinct biological functions: the Queen, the Drones and the Workers bees [Winston, 1991]. The figure bellow presents these three kinds of bees.

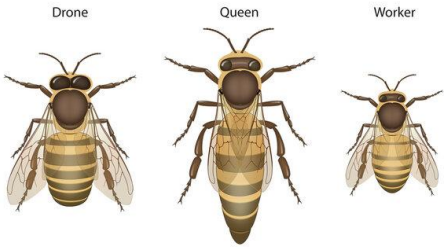


Figure 4.1: Kind of bees. (image from https://www.britannica.com/animal/honeybee, 06/20/2025)

2.1. The Queen Bee: Heart of the Hive's Reproduction

The queen bee holds the vital role of reproduction as the sole egg-layer in the colony, ensuring its survival and growth. Her biological functions are finely tuned for this purpose. She regulates the entire colony's behavior through pheromones, which suppress worker bees' ovary development and maintain social cohesion. During a single mating flight early in her life, she mates with multiple drones and stores their sperm, enabling her to fertilize eggs throughout her lifespan, typically two to five years. With a remarkable capacity to lay up to 2,000 eggs per day, she directly controls the colony's population dynamics and genetic diversity. This prolific reproduction is essential for sustaining the hive's workforce, replacing aging bees, and facilitating colony expansion through swarming. The queen's health and productivity are so critical that worker bees will replace her if her egg-laying declines, demonstrating the colony's intricate balance between individual specialization and collective survival.

2.2. Drones: The Transient Males of the Hive

Drones serve one critical function: mating with virgin queens to ensure genetic diversity. Unlike workers, they lack stingers and foraging ability, depending entirely on the colony for sustenance. Their reproductive role ends dramatically - during mating, their genitalia detach (a phenomenon called mating trauma), resulting in immediate death. As winter approaches, workers expel surviving drones to conserve resources, demonstrating the colony's ruthless efficiency in energy allocation.

2. 3. Worker Bees: The Industrious Backbone of the Hive

Worker bees, the sterile females of the colony, perform all labor-intensive tasks necessary to sustain the hive. Their roles shift in a precise age-based progression known as

temporal polyethism. In their first days (1-12), they serve as nurse bees, diligently feeding larvae with nutrient-rich royal jelly, tending to the queen's needs, and maintaining hive hygiene by cleaning brood cells. Between days 12 and 18 of their development, they take on the role of hive builders, producing wax to build honeycombs and carefully storing the gathered nectar and pollen. Approaching adulthood (days 18-21), they take on guard duties, aggressively defending the hive entrance from predators by releasing alarm pheromones and deploying their stingers when necessary. Finally, in the last stage of their lives (day 21+), they become foragers, embarking on daily expeditions up to 5 kilometers from the hive to gather nectar, pollen and water. These experienced bees communicate complex information about food sources through their intricate waggle dances, enabling efficient resource collection for the entire colony. This remarkable division of labor ensures optimal hive functioning, with each bee contributing specialized skills at just the right time in their development.

2. 4. Biological Roles of Workers Bees in Colony Foraging

The minimal model of a honey bee colony comprises three groups: employed bees, onlooker bees, and scout bees. Employed bees explore food sources and share information with onlooker bees, which then evaluate and select food sources based on this information. Higher-quality food sources have a greater probability of being chosen by onlooker bees, while lower-quality ones are more likely to be abandoned. If an employed bee's food source is rejected due to poor quality, it transitions into a scout bee and begins searching randomly for new food sources.

Bees communicate food source information to their swarm through distinct dance forms:

1. Round Dance: Performed when the food source is close to the hive.
2. Waggle Dance: Used for distant food sources; the speed of the dance indicates the distance (faster mean closer).
3. Tremble Dance: Signals that the bee is struggling to unload nectar and lacks current knowledge of the food source’s profitability.

Thus, exploitation is carried out by employed and onlooker bees, whereas exploration is handled by scout bees. The next sub-section details how these bee-inspired mechanisms are implemented in the Artificial Bee Colony (ABC) algorithm.

3. Artificial Bee Colony (ABC) Algorithm

3.1. Description of the algorithm

The ABC algorithm is an evolutionary algorithm bio-inspired [Karaboga & Basturk, 2007]. It imitates the honey bee swarms in food foraging. It assumes the existence of a set of operations that may resemble some features of the honey bee behavior. For instance, each solution within the search space includes a parameter set representing food source locations. The “fitness value” refers to the food source quality that is strongly linked to the food’s location. The process mimics the bee’s search for valuable food sources yielding an analogous

process for finding the optimal solution. It operates through the collaboration of three types of bees: employed bees, onlooker bees, and scout bees, each with distinct roles in the search for nectar (or optimal solutions).

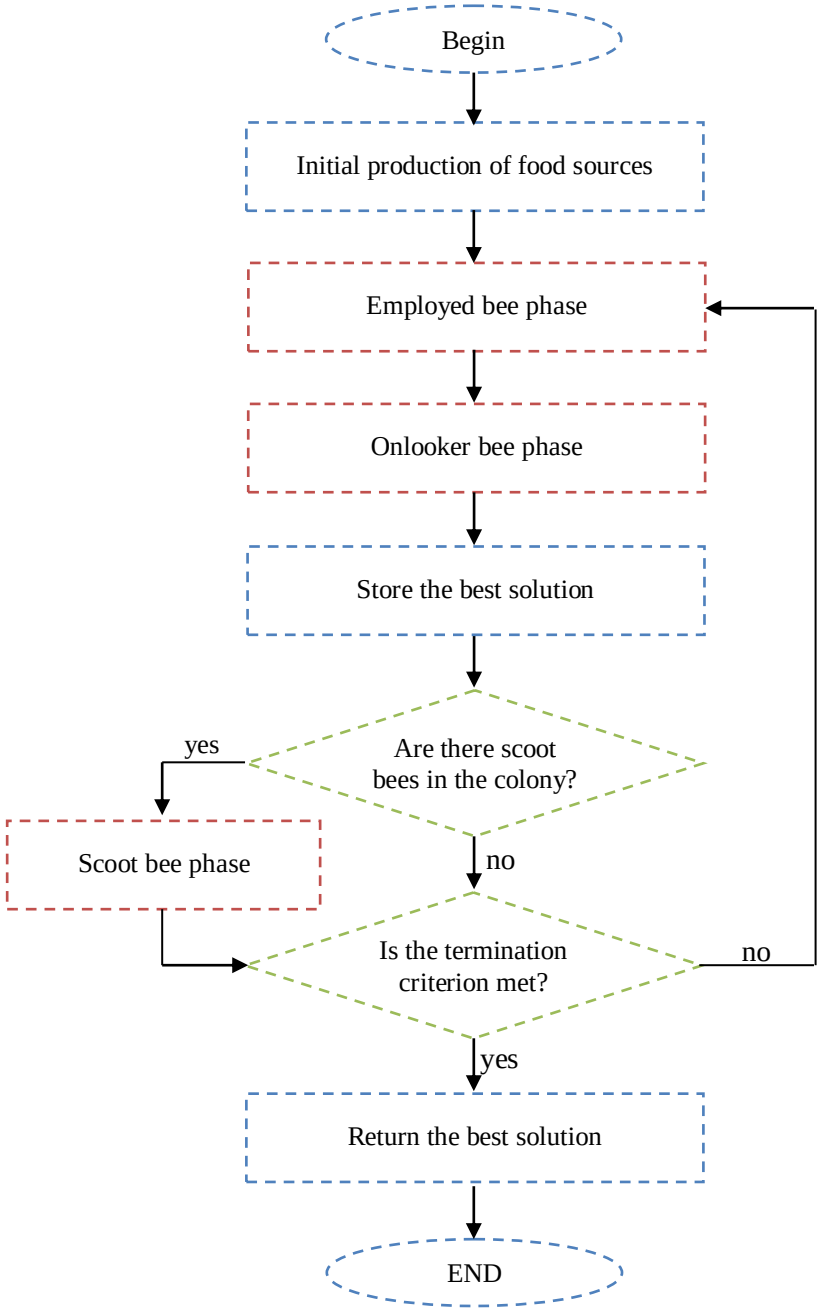


Figure 4.2: Flowchart of ABC algorithm.

The employed bees are responsible for exploiting known food sources. Each employed bee represents a potential solution and assesses its quality based on a fitness function. They search in the vicinity of their assigned food source and can adjust their position to improve the

solution. If a bee finds a better solution, it shares this information with the onlooker bees. The later monitor the quality of food sources shared by employed bees. They utilize a probability-based selection mechanism to choose which food source to explore based on its fitness. By concentrating on the most promising sources, onlooker bees contribute to the exploitation phase of the algorithm called also local search, further refining the search for optimal solutions. The scout bees present the explorative phase and they are responsible for exploring new areas of the search space to discover new food sources.

Their random search helps maintain diversity in the population and prevents the algorithm from getting trapped in local optima. Through the coordinated efforts of these three types of bees, the ABC algorithm efficiently explores and exploits the solution space. Figure 4.1 presents the flowchart of the ABC algorithm.

3.2. ABC Algorithm operation

The ABC algorithm operates throws a set of steps. It begins food foraging (solution search) by producing randomly an initial population of *NS (number)* food sources (bees) in search space which are uniformly distributed between the pre-specified lower and upper values.

Each bee is created according to the equation (4.1):

$$b_i = b_{min} + rand(0,1) * (b_{max} - b_{min}) \qquad i = 1, \dots, NS \tag{4.1}$$

where b_i is a bee, b_{min} and b_{max} are the upper and the lower values of the search space respectively.

Additionally, a counter that tracks the number of solution attempts is reset to zero for each bee in this phase. After initialization, the initial food sources (solutions) undergo repeated improvement cycles, where employed bees and onlooker bees explore the neighborhood and enhance the solutions.

After the initialization phase, the ABC algorithm evaluates the initial population and performs the three following steps until convergence to the optimal global solution (satisfactory fitness) or maximum iterations.

Step 1: Send Employed Bees (Employed bee phase)

There are as many employed bees as there are food sources. In this stage, each employed bee creates a new food source in the vicinity of its current position using the following equation:

$$v_i = b_i + \varphi_i(b_i - b_k) \qquad i = 1, \dots, NS \tag{4.2}$$

where the scale factor φ_i is uniformly distributed random number between $[-1,1]$, b_i and b_k are the i^{th} food source (solution) and one of *NS* food source in the vicinity respectively ($i \neq k$). v_i represents the new generated food source (solution).

Once a new solution v_i is generated, its fitness value (nectar amount), representing the solution's profitability, is calculated. The fitness value of a solution v_i can be represented by the objective function or computed using the following equation:

$$F(b_i) = \begin{cases} \dfrac{1}{1+J_i} & if\ J_i \geq 0 \\ 1 + abs(J_i) & if\ J_i < 0 \end{cases} \tag{4.3}$$

J_i represents the objective function to be optimized.

A greedy selection is applied between v_i and b_i. If v_i has a better fitness, b_i is replaced by v_i; otherwise, b_i is retained

Step 2: Select the Food Sources by the Onlooker Bees (Onlooker bee phase)

Each onlooker bee (the number of onlooker bees corresponds to the food source number) selects a food source b_i with a probability proportionally to the nectar amount. The probability P_i that the food source b_i will be selected is calculated according to the following equation:

$$\boldsymbol{P_i = \frac{f(b_i)}{\sum_{j=1}^{NS} f(b_j)} \qquad i = 1, \dots, NS} \tag{4.4}$$

where $f(b_i)$ is the fitness of the solution b_i.

The probability of a food source being selected by onlooker bees increases with an increase in the fitness value of the food source. Upon selection, onlooker bees visit the food source and generate a new candidate position within its vicinity, as defined by equation (4.2). If the new solution's fitness (nectar amount) improves upon the previous value, the position is updated. Otherwise, the original solution is kept.

Step 3: Determine the Scout Bees (Scout bee phase)

When a food source (a candidate solution) fails to improve after a set number of trials (called the *limit*), it is considered exhausted. The employed bee abandon the food source associated with it and becomes a *scout* bee. Scout bees explore the search space without prior knowledge, generating a new solution randomly using equation (4.1).

To track whether a candidate solution has reached the *limit*, each bee b_i associated with this food source has a counter. This counter increases each time a bee fails to enhance the food source's fitness.

Occasionally, scout bees may discover highly promising, previously unknown food sources by chance.

Step 4: Termination

If the stopping criterion is met or the maximal iteration number is reached, return the best bee (optimal solution).

To reach the global optimum, the ABC Algorithm balance between exploitative search and exploratory search and the both in random manner.

Algorithm 1 bellow describes the ABC algorithm

Algorithm 1: ABC algorithm

1: Define *SN* (population size), b_{min} (lower value) and b_{max} (upper value), *limit* (control parameter), MaxIt (maximum iteration number)
2: Generate randomly *SN* bees (b_i, i=*1*...*SN*) in the search space to form an initial population using equation (4.1)
3: Evaluate the fitness function of all the bees $f(b_i)$
4: Keep the best bee (best solution in the population) b_{best}
5: For each bee b_i, fix *"no-improvement-cycle$_i$"* to *0*.
6: Set Iteration=1
7: Generate a candidate solution v_i for each bee b_i by equation (4.2)
8: Evaluate the fitness function of all the candidate solutions $f(v_i)$
9: If $f(v_i)$ is better than $f(b_i)$, $b_i = v_i$, set *no-improvement-cycle$_i$* to 0; otherwise increment *no-improvement-cycle$_i$*
10: Calculate the probability values P_i by equation (4.4)
11: For each bee b_i selected depending on its probability P_i, generate a new solution v_i
12: Evaluate the fitness function of all the new solutions $f(v_i)$
13: If $f(v_i)$ is better than $f(b_i)$, $b_i = v_i$, set *no-improvement-cycle$_i$* to 0; otherwise increment *no-improvement-cycle$_i$*
14: For each bee b_i if *no-improvement-cycle$_i$* > *limit* generate a new solution for b_i according to equation (4.1).
15: Keep the best bee (best solution in the population) b_{best}
16: iteration =iteration +1
17: If iteration > MaxIt return the best solution achieved so far (b_{best}) and Stop, otherwise go to 7.

4. Hybrid FCM-ABC Method for Medical Image Segmentation

In this section, a new enhancement of FCM called Hybrid FCM-ABC method is introduced; it is based on the ABC Algorithm [Mokhtari et al., 2025]. Although the FCM has advantages like efficacy, simplicity and computational efficiency, it nonetheless has major drawbacks such as number of clusters, cluster centers values and is easily trapped in local optima. So, the main objective is to overcome these major drawbacks that will affect the clustering in term of precision. For this purpose, we improve the FCM clustering by exploiting ABC algorithm in order to find simultaneously the right number of clusters and the optimal clusters centers for a given image *I* of *N* pixels. The right values of these parameters

leads to the optimal solution. ABC algorithm combines between exploitation and exploration to find the optimal values of FCM parameters. It ensures the searching in all directions in the solution space.

The problem is regarded as follow. From the hive, the center of the search space, bees search food source in different areas around it. The areas present subspaces of search. Each subspace of search is far from the hive with a distance that represents the number of clusters. Subspace with distance 2 from the hive contains all bees (probable solutions) having 2 cluster centers values and subspace with distance *k* from the hive contains all bees (probable solutions) having *k* cluster centers values (cf.fig.4.3). The quantity of food in each emplacement within subspace represents cluster centers values.

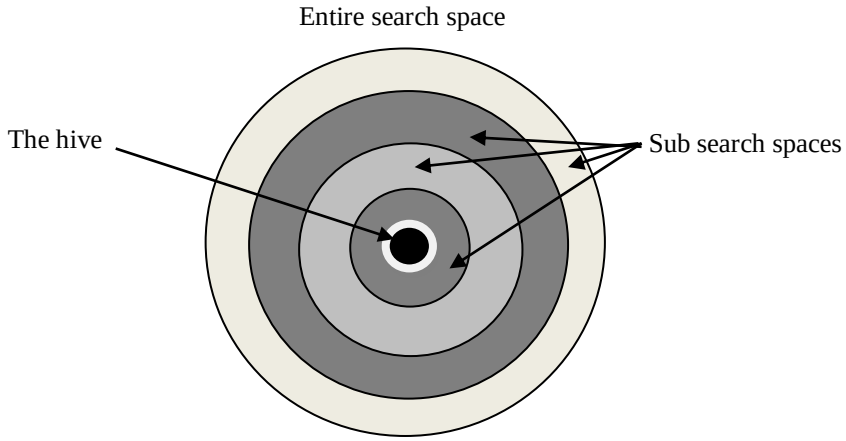


Figure 4.3: Decomposition of search space

In the beginning of the food foraging in search space, all sub space has the same probabilities to get the best food source. However, the bees are distributed in the entire search space. After communicating information about source food, the majority of bees concentrate in the most promising subspaces search and so on. Therefore, discovering of the most food-rich location will be very likely. Figure 4.4 illustrates this behavior.

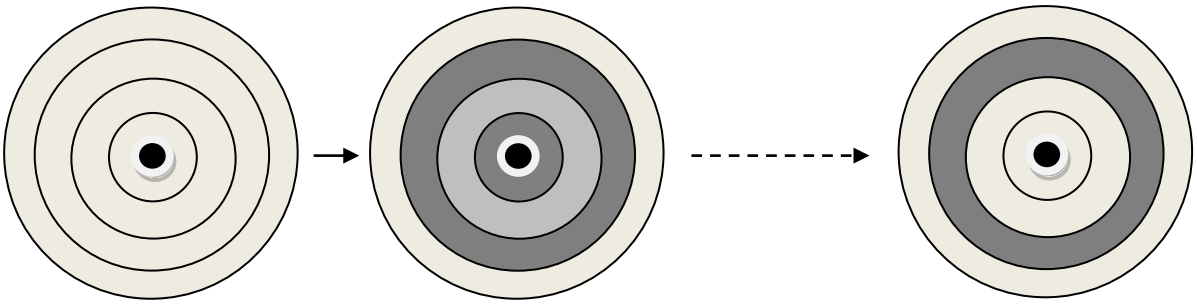


Figure 4.4: Searching behavior (dark zone around the hive contains most promising bees)

4.1. Data Structure

To achieve the objective of finding the optimal solution, each bee b_i in the population is represented as a vector composed of two distinct parts:

1. The first part encodes the *number of clusters*, determining the structure of the solution.
2. The second part maintains *the values of the cluster centers*, defining their positions in the search space.

This dual-vector representation ensures that both the *number of clusters* and their values are optimized simultaneously (cf. Fig. 4.5 for a visual illustration).

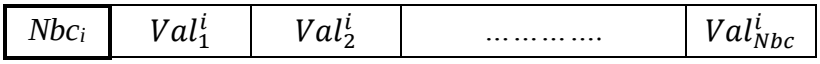


Figure 4.5: Data structure of artificial bee.

where Nbc_i is the number of clusters of the image to be segmented. This number is between *2* and maximum number of cluster (*MaxNbc)*. The value *MaxNbc* depends on biological context. For example, *MaxNbc* =7 is sufficient for brain image. Val_j^i is the value of the center c_i of the bee b_i which is the grey levels of the input image *I*. These values are in the range *[0, 255]*. Figure 4.6 bellow presents four bees with different number of clusters with different values. The first bee encodes three clusters centers with their values. The second encodes also three clusters centers with their values while the third and the fourth bee encode two and five clusters centers respectively with their values.

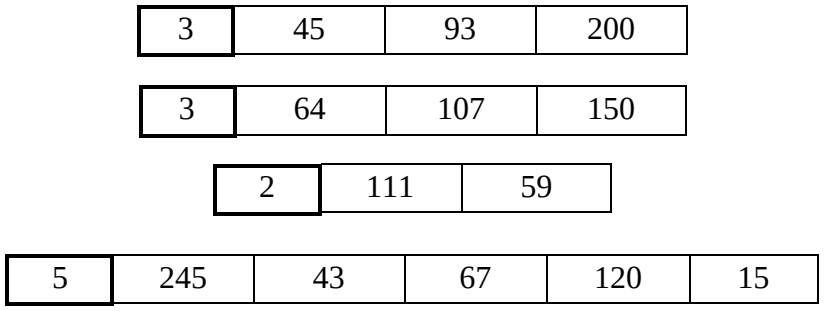


Figure 4.6: Four examples of artificial bees with different configuration

4.2. Objective Function

To evaluate the quality of the food, number and values of cluster centers (solution), we have developed a novel composite objective function *J* designed to simultaneously optimize both cluster quality (cluster values) and model complexity (number of clusters). This function addresses two fundamental challenges in unsupervised clustering:

1. Geometric optimization of cluster centers (minimizing intra-cluster dispersion),
2. Automatic determination of the optimal number of cluster.

Function *F* combines the conventional Fuzzy C-Means (FCM) objective function, preserving the algorithm's ability to find spatially coherent clusters and a cluster validity index component that penalizes solutions with either too many or too few clusters.

The combined objective function is formally defined as follows:

$$\boldsymbol{J(b_i) = W_1 F1(b_i) + W_2 F2(b_i) \qquad i = 1, \dots, NS} \tag{4.5}$$

where $F1(b_i)$ corresponds to the standard Fuzzy C-Means (FCM) objective function, which minimizes the weighted sum of squared distances between data points and cluster centers. The second term, $F2(b_i)$, represents a clustering validity index that evaluates the quality of the resulting partitions in terms of compactness and separation. The weights W_1 and W_2 control the relative importance of each component in the overall optimization process. *NS* is the population size: number of bees in search space.

The motivation behind this hybrid formulation lies in addressing the limitations of using FCM alone. While FCM effectively minimizes intra-cluster variance, it does not inherently ensure well-separated or meaningful clusters, especially when the optimal number of clusters is unknown or the data contains overlapping structures. Incorporating a validity index as an additional criterion enhances the ability of the algorithm to identify more compact and distinct clusters, thereby improving overall segmentation quality.

By combining both objectives, the proposed function enables a balanced trade-off between minimizing within-cluster distortion (via FCM) and maximizing cluster validity (via the validity index). This dual-objective approach proves particularly beneficial in complex applications such as medical image segmentation, where accurate and interpretable clustering is essential for diagnostic reliability.

Both weights W_1 and W_2 can be adjusted depending on the specific requirements of the application or based on prior knowledge about the data structure.

According to the structure of bee b_i, $F1$ is defined as:

$$F1(b_i) = \sum_{k=1}^{Nbc_i} \sum_{j=1}^{N} u_{k,j}^{m} d^2(x_j, Val_k^i) \tag{4.6}$$

x_j is the image pixels and *d* is the Euclidean distance.

F2 is a cluster validity index, known as the IMI (IMbalanced Index) index, proposed by Yun Liu [Liu et al., 2021b] to identify the optimal number of clusters. It is formally defined in equation (4.7). IMI allows us to solve a key challenge in clustering analysis which is determining the optimal number of clusters in a dataset. According to the experimentations deal on different data set, authors in [Liu et al., 2021b] confirm the robustness of IMI. Thereby, it is a well candidate to be used as strong tool to determinate the right number of clusters.

$$F2(b_i) = \frac{\sum_{k=1}^{Nbc_i} \frac{\sum_{j=1}^{N} u_{k,j}^m d^2\left(x_j, Val_k^i\right)}{\sum_{j=1}^{N} u_{k,j}^2}}{\min\limits_{l \neq k} \delta_{l,k} d^2\left(Val_l^i, Val_k^i\right) + \operatorname{median}\limits_{l \neq k} \delta_{l,i} d^2\left(Val_l^i, Val_k^i\right)} \tag{4.7}$$

where $\delta_{l,k} = \frac{\sum_{j=1}^{N} u_{l,j}}{\sum_{j=1}^{N} u_{k,j}}$.

4.3. General steps of the Hybrid FCM-ABC Method

The general steps of the Hybrid FCM-ABC Method are outlined as follows. These steps integrate the strengths of the FCM algorithm and the ABC optimization technique to achieve robust and accurate segmentation results:

Step 1- Initialization: we set the maximum number of clusters *MaxNbc,* length of the worst bees *L*, the number of trial *limit*, a maximum number of iteration *MaxIteration* and a threshold ϵ. Then an initial population of *NS* bees is generated in which each bee b_i, in its first part ought to be assigned a random value in the range *[2, MaxNbc]*. According to the grey levels of the image *I*, each value Val_j^i in second part has a value in the range *[0, 255]*. It is initialized using equation (4.1). For each bee b_i, we set the counter *"no-improvement-cycle"* to *0*.

Step 2- Fitness evaluation: since each bee b_i encodes cluster centers values, we calculate the membership value $u_{kj}^i$for each cluster centers c_k^i for all the bees (b_i, *i=1,…, NS*) using equation (4.8) bellow.

$$u_{kj} = \frac{(d^2(x_j, Val_k^i))^{\frac{1}{1-m}}}{\sum_{l=1}^{Nbc_i} (d^2(x_j, Val_l^i))^{\frac{1}{1-m}}} \tag{4.8}$$

Then, we evaluate the objective function of the bees in the population $J(b_i)$, according to the equation (4.5) and their fitness with equation (4.3) .The bee with the best configuration is stored (b_{best}).

Step 3- Employed Bee Phase: in this step, each employed bee generates a new solution in its neighborhood according to the equation (4.2). It consists of modifying slightly each center c_k^i of each bee b_i to find a better position through local exploration without affecting the number of clusters Nbc_i. Then, new solution's fitness is evaluated. If the new solution is better, replace the current bee by this new solution. Otherwise increment the counter *"no-improvement-cycle"*.

Step 4- Onlooker Bee Phase: based on the fitness values, we assign probability P_i to each solution b_i using the equation (4.4). We generate randomly a number r in the range *[0, 1]*. If r is less then P_i, each onlooker applies modifications on b_i using the equation (4.2) to further refine the clusters centers.

Step 5- Scout Bee Phase: to enhance the capability to exploit the global search, we sort the bees according to equation (4.3) and we abandon all bees that the "*no-improvement-cycle*" exceeds *limit*. If any abandoned bee belongs to the list of the L highest bees, we replace the abandoned bees with new configurations in which we keep the number of clusters and we reset only the cluster centers with equation 4.1. If any abandoned bee doesn't belong to this list it will be reinitialized completely with a random number of clusters in the range *[2, MaxNbc]* and new cluster centers values using equation (4.1).

Step 6- Loop: steps from 2 to 5 are repeated until the objective function J became less than the threshold ϵ or the maximum number of iterations *MaxIteration* is reached.

Step 7- **Termination**: finally, we use the best configuration stored so far (b_{best}). The number of clusters and their centers values encoded in its configuration are used to perform a last calculation of pixel memberships u_{kj}^{best} according to equation (4.8). We assign each pixel x_i of the image I to center for which the memberships u_{kj}^{best} is higher for the purpose to generate the segmented image.

4.4. Hybrid FCM-ABC Algorithm

Our proposed hybrid FCM-ABC method is summarized in the flowchart presented below (cf.fig.4.7). The flowchart outlines the key steps and logic of the hybrid FCM-ABC method, highlighting how the ABC algorithm is integrated with the FCM framework to achieve robust and accurate segmentation results. Each step corresponds to a specific phase of the optimization process, ensuring clarity and reproducibility of the method.

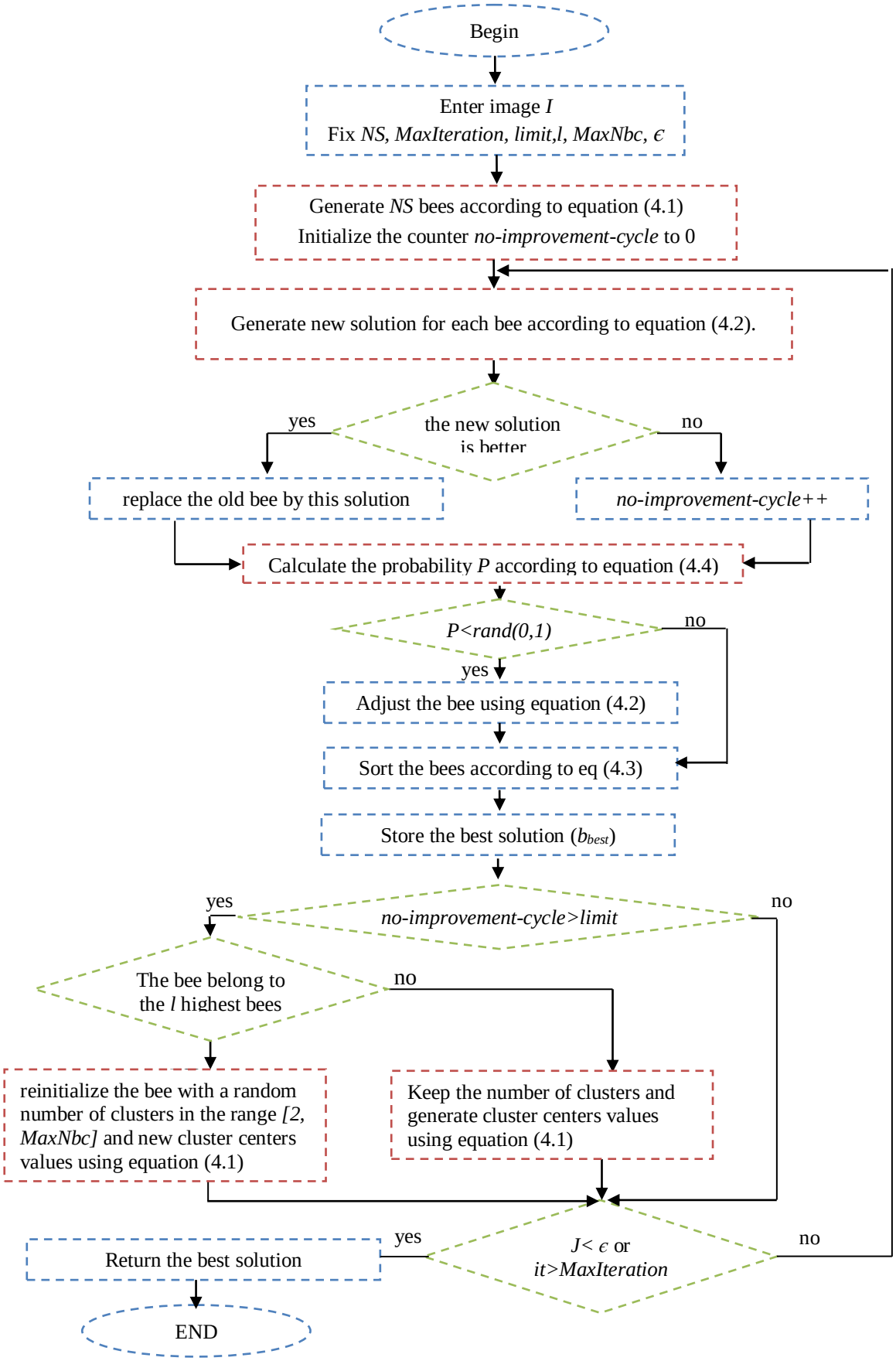


Figure 4.7: Flowchart of Hybrid ABC-FCM method.

Algorithm 2 presents the pseudo code of our proposition.

Algorithm 2: Hybrid FCM-ABC algorithm

Input: original image *I*
fix the parameters: *MaxNbc, NS, ϵ, limit, L, MaxIteration.*
generate randomly an initial population of bees b_i $(i = 1, 2, \dots, NS)$ using equation (4.1).
it=0
for each bee b_i, fix *"no-improvement-cycle$_i$"* to *0*
repeat
it=*it*+1
for each bee b_i
    calculate the membership value u_{ij} using (4.8)
    calculate the fitness function $F(b_i)$ according to the equation (4.3).
endfor
Bbest = bee b_i with the highest fitness *Fl*.
for each bee b_i
    generate a new solution *bnew* according to the equation (4.2).
    evaluate the *bnew*'s fitness.
        If *bnew* is better, $\boldsymbol{b_i}$=*bnew*.
        else *"no-improvement-cycle"*++.
    calculate the solution probability P_i using the equation (4.4).
endfor
for each bee b_i
    generate a random number *r* in the range *[0, 1]*
    if P_i< *r*, update b_i with equation (4.2)
    evaluate its fitness according to equation (4.3).
ElitBee =L highest bees in term of fitness
for each bee b_i
    if *"no-improvement-cycle" > limit*
        if $\boldsymbol{b_i} \in \boldsymbol{ElitBee}$, replace $\boldsymbol{b_i}$ with new clusters centers according to equation (4.1) without affecting the number of clusters *Nbc$_i$*
        else generate a new solution for $\boldsymbol{b_i}$ according to equation (4.1).
    reset to zero the counter *"no-improvement-cycle"*
Endfor
until (*Fl <ϵ or it>=MaxIteration)*
calculate the membership value u_{ij} according to *Bbest*.
cluster pixels of the image *I* according to the membership value u_{ij}

5. Experimental Results

5.1. Experimental setup

The performance of the Hybrid FCM-ABC method depends on several key parameters. These parameters are selected to balance exploration, exploitation, and computational efficiency.

The population size refers to the total number of bees, including employed, onlooker, and scout bees, typically set between 50 and 100. This range balances exploration and computational efficiency: a larger population enhances solution diversity and search space exploration, helping avoid local optima, while a smaller size reduces computational overhead. In our case, for brain MRI segmentation, a population size of 50 is chosen as it effectively explores the high-dimensional search space of cluster centers without incurring excessive computational costs.

In all our implementation, the maximum number of iterations is set to 300. Typically, values between 100 and 500 iterations are recommended in optimization tasks, including medical image segmentation. The number of iterations plays a crucial role in balancing exploration and computational efficiency, the higher the number, the more thoroughly the algorithm can explore the search space and refine potential solutions. However, this also results in increased computation time. In our context of brain MRI segmentation, where convergence is often achieved within this range, 300 iterations provide a reasonable trade-off between accuracy and performance, allowing the algorithm to converge effectively without unnecessary resource consumption.

To avoid stagnation in a local minimum, we set maximum number of cycles (*limit*) to 10, which limits the number of consecutive cycles without improvement and helps maintain a balance between exploration and exploitation during the optimization process.

A limit of 10 cluster centers (*MaxNbc*) was imposed, based on empirical observation that no brain MRI image in the analyzed dataset exhibited a greater number of distinct segments.

In the objective function, the weights W_1 and W_2 are both set to 0.5, ensuring a balanced contribution of the individual components in the optimization process. And finally, the size of the list *l* is set to 20.

5.2. Metrics used for segmentation evaluation

The evaluation of brain MRI segmentation performance relies on several metrics to quantify accuracy, robustness, and consistency [Taha & Hanbury, 2015]. In cases where the ground truth is available, we use Jaccard Similarity Index. In cases where the ground truth is unavailable, it becomes necessary to rely on internal validation indices to evaluate the quality of the clustering results such as Partition Coefficient Index, Partition Entropy Index and Davies-Bouldin Index. By utilizing these indices in combination, we can obtain a comprehensive evaluation of the clustering outcomes, ensuring that the proposed method

achieves optimal performance even in the absence of ground truth information. This approach not only enhances the reliability of the segmentation process but also enables meaningful comparisons with other clustering techniques under similar conditions.

5.2.1. Jaccard Similarity Index

The Jaccard Similarity Index (also called Jaccard Similarity Coefficient or Intersection over Union, IoU) is a statistical measure used to evaluate the similarity between two sets. In the context of medical image segmentation, it quantifies the overlap between a predicted segmentation and the ground truth (manual annotation).

The Jaccard Similarity Index (*JSI*) ranges from 0 to 1, where 0 indicates no spatial overlap between the predicted segmentation and the ground truth, and 1 denotes a perfect match. Higher JS values reflect greater segmentation accuracy, as they signify stronger agreement between the automated output and the reference standard. This metric is particularly useful for quantifying volumetric overlap in tasks such as tumor or lesion segmentation in medical imaging like MRI or CT image, where precise boundary delineation is critical. The *JS* for a cluster *k* is defined as:

$$JSI_k = \frac{|A_k \cap B_k|}{|A_k \cup B_k|} \tag{4.9}$$

where A_k and B_k are the total number of pixels labeled into the cluster *k* identified by the clustering algorithm and the ground truth respectively. The cluster *k* is well detected when the value of JSI_k is near 1.

5.2.2. Partition Coefficient Index

The Partition Coefficient Index (*PCI*), also known as the fuzzy partition coefficient, is a metric used to evaluate the quality of fuzzy clustering algorithms, such as the Fuzzy C-Means (FCM) method. Unlike crisp clustering, where each data point belongs exclusively to one cluster, fuzzy clustering assigns membership degrees, indicating how strongly a point is associated with each cluster. The PCI quantifies the fuzziness of the resulting partition by measuring the average squared membership values across all data points and clusters.

The PCI is widely used in medical image segmentation, particularly in algorithms that handle uncertainty, such as brain components delineation in MRI image. By optimizing clustering algorithms to maximize *PCI*, we can improve the reliability of automated segmentation results. The *PCI* is defined as:

$$PCI = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^2 \tag{4.10}$$

The *PCI* ranges from $\frac{1}{C}$ to 1, where *C* represents the number of clusters. A value of $\frac{1}{C}$ corresponds to a completely fuzzy partition, indicating no meaningful clustering structure (uniform membership distribution across all clusters). Conversely, a *PCI* of 1 signifies a perfectly crisp partition, where each data point unequivocally belongs to a single cluster. Higher *PCI* values reflect reduced fuzziness and sharper separation between clusters, suggesting improved clustering quality.

5.2.3. Partition Entropy Index

The Partition Entropy Index (*PEI*) is a widely used metric for evaluating the fuzziness and uncertainty in fuzzy clustering algorithms, such as FCM. Unlike *PCI*, which measures the crispness of clustering, *PEI* quantifies the degree of disorder or uncertainty in the membership assignments of data points across clusters. The *PEI* is defined as:

$$PEI = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}\, log(u_{ij}) \tag{4.11}$$

PEI exhibits a theoretical range from 0 to log(C), where C represents the number of clusters. A PEI value of 0 indicates perfectly crisp clustering. Conversely, the upper bound of log(C) corresponds to maximal fuzziness, occurring when membership degrees are uniformly distributed across all clusters. In practical applications, lower *PEI* values signify more definitive cluster assignments and reduced uncertainty, while higher values reflect increased ambiguity in the partitioning. This inverse relationship between *PEI* values and partition certainty makes it particularly valuable for assessing the reliability of fuzzy clustering algorithms in applications such as medical image segmentation, where uncertainty in tissue classification (tumor vs. healthy tissue in MRI) must be quantified. It is also applied in pattern recognition and bioinformatics to assess the reliability of fuzzy clustering results.

5.2.4. Davies-Bouldin Index

The Davies-Bouldin Index (*DBI*) is a metric for evaluating clustering algorithm performance by quantifying the trade-off between intra-cluster compactness and inter-cluster separation. Unlike validity measures that assess fuzzy partitions (like *PCI* or *PEI*), the *DBI* is specifically designed for crisp clustering solutions. In the case of soft clustering, it is used after defuzzification step of data affectation. The *DBI* is defined as:

$$DBI = \frac{1}{C}\sum_{i=1}^{C} \underset{i \neq j}{max}\left(\frac{S_i + S_j}{D_{i,j}}\right) \tag{4.12}$$

Where S_i is the mean distance between the center of the cluster *I* and all the points belonging to this cluster and $D_{i,j}$ denotes the distance between the centroids of the clusters *I* and *J*.

The *DBI* produces strictly non-negative values, with lower values indicating superior clustering quality. The index approaches its theoretical optimum of zero when clusters exhibit both high intra-cluster compactness and strong inter-cluster separation. As a comparative metric, the *DBI* should be minimized when evaluating alternative clustering solutions. The index offers three key advantages: (1) its intuitive interpretation directly captures the separation-to-compactness ratio, providing immediate insights into cluster validity; (2) inherent scale invariance ensures consistent performance across differently scaled datasets, as distances are normalized relative to cluster dispersion; and (3) computational efficiency, requiring only centroid positions and dispersion measures. These characteristics make the *DBI* particularly valuable for medical image analysis, where rapid evaluation of tissue segmentation quality is often required.

5.3. Experimental results on Simulated Brain MR Images

The following experiments were conducted using Simulated Brain Database (SBD[1]). The SBD provides synthetic MRI brain images with known ground truth segmentations, making it ideal for validating segmentation algorithms. The images simulate different intensity inhomogeneities, and slice thicknesses, mimicking real-world MRI challenges. This database includes ground truth information for tissue of white matter (WM), and grey matter (GM), and cerebrospinal fluid spaces (CSF). It offers a controlled setting to assess the algorithm's accuracy and its ability to handle intensity inhomogeneity effectively.

The proposed Hybrid FCM-ABC method was initially tested on a T1-weighted brain MRI images with dimensions of 217×181 pixels, which includes 20% grayscale non-uniformity to simulate real-world imaging challenges (cf.fig.4.8). The main objective of our proposed method is to accurately segment and identifies critical brain regions, namely WM, GM and CSF. These tissue types are fundamental for radiologists in their analysis and diagnosis of various neurological disorders and diseases. Figure 4.8 shows three T1-weighted brain MRI images in X87, X94 and X105 planes.

5.3.1. Clusters number detecting

First, we will show the ability of our method to find the right number of cluster centers and how to converge to this right number across iterations.

Several images are tested. For illustrating this outcome, we have chosen two T1 weighted images in X94 and X105 planes. Both images have four clusters namely WM, GM, CSF and the background.

The presented figures 4.9 and 4.10 illustrate the evolution of a clustering performance metric across increasing computational cycles (from 0 to 300, in increments of 30) for cluster numbers ranging from 2 to 10. A critical analysis of the both images strongly suggests that 4 clusters represent the optimal partitioning, as evidenced by its superior and sustained performance over alternative cluster counts. This conclusion is supported by three key

[1] Brain Web: Simulated Brain Database, http://brainweb.bic.mni.mcgill.ca/brainweb/, accessed 20 September 2024.

observations: performance dominance, robustness over iterations, and resistance to overfitting.

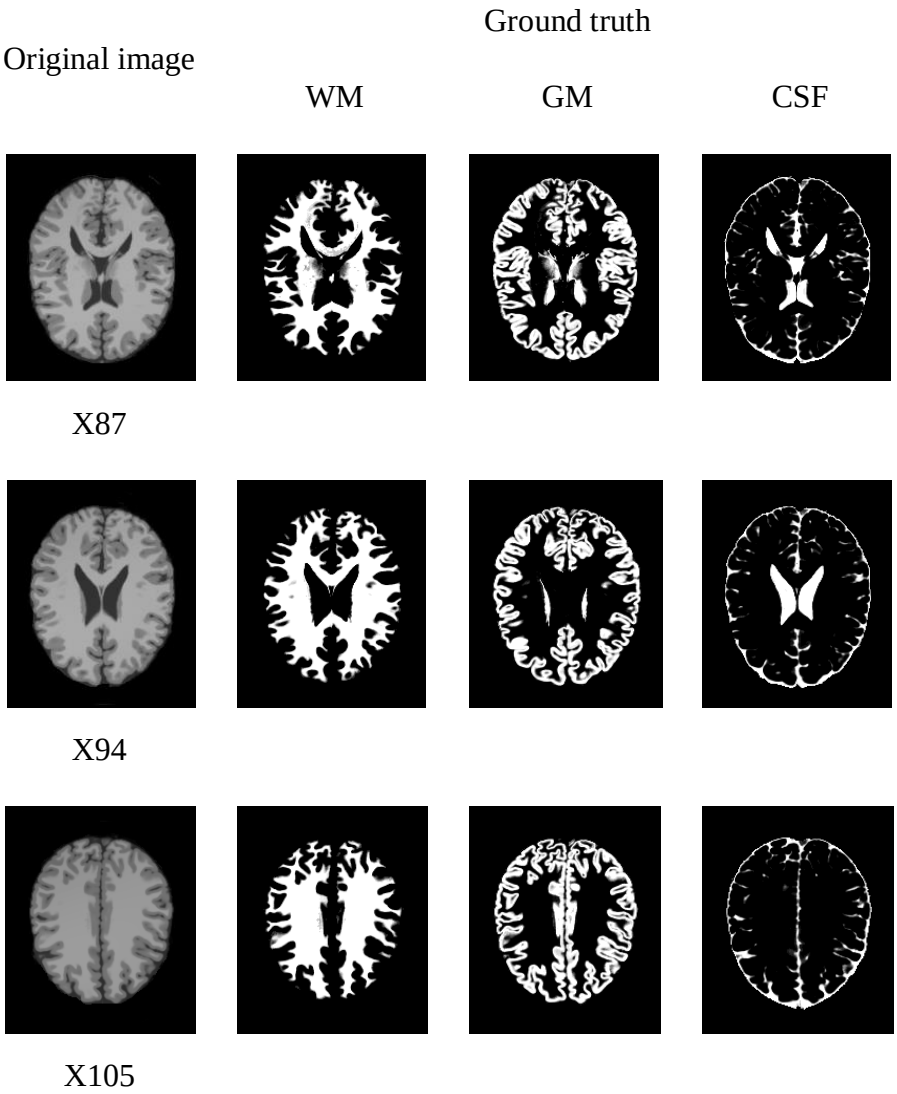


Figure 4.8: Brain MRI images in X87, X94 and X105 planes with their ground truth

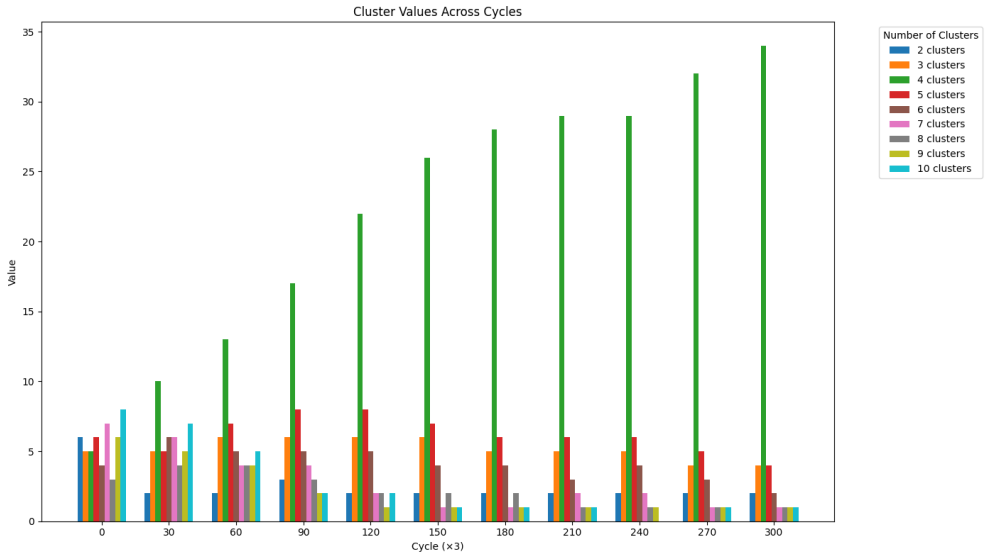


Figure 4.9: Evolution of clusters number across iterations for T1-weighted brain MRI image in X94 plane.

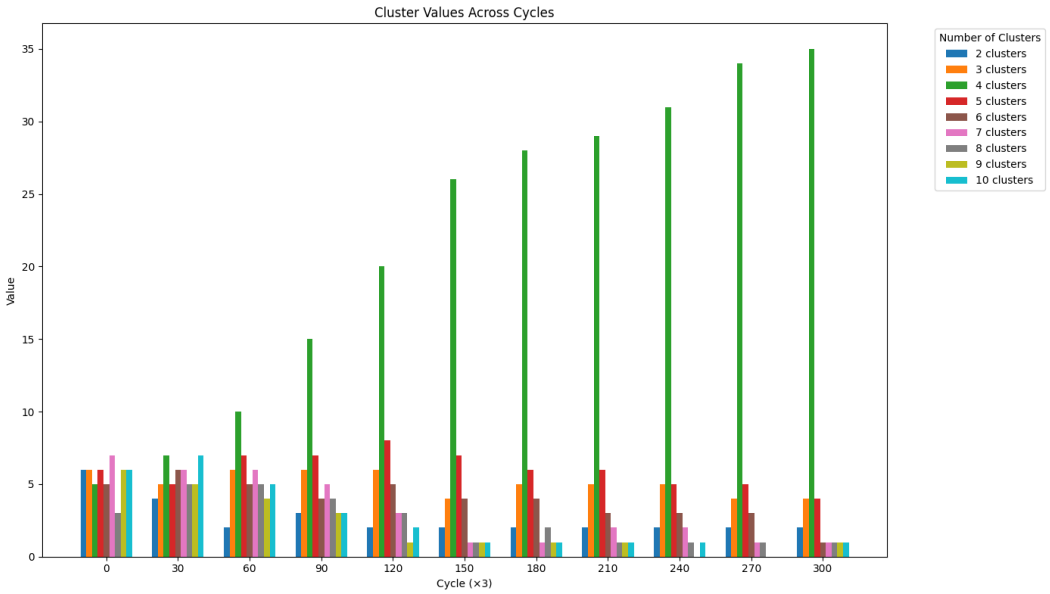


Figure 4.10: Evolution of clusters number across iterations for T1-weighted brain MRI image in X105 plane.

First, the performance dominance of 4 clusters is unequivocal. While all cluster counts begin with relatively comparable metrics at the initial cycle (5 bees for 4 clusters vs. 6 bees for 2 clusters and 7 bees for 7 clusters), the trajectory of 4 clusters diverges markedly as computational iteration progress. By the 300th iteration, the metric for 4 clusters peaks at 34 bees, which is not only the highest value across all cluster counts but also around 8.5 times greater than its nearest competitor. This substantial gap underscores the capacity of 4 clusters to better capture the underlying structure of the images.

Second, the robustness of 4 clusters is demonstrated by its monotonic improvement over cycles. The metric for 4 clusters exhibits a steady and uninterrupted rise from 5 (at 0 cycles) to 34 (at 300 iterations), indicating that additional computational iterations refine and enhance the clustering solution. This behavior contrasts sharply with other cluster counts, such as 5 or 6, which plateau or decline after an initial period of competitiveness (5 clusters peaks at 8 bees by 90 cycles before regressing to 4 bees). Such trends suggest that these cluster counts may initially approximate the data structure but ultimately fail to generalize effectively as the algorithm iterates.

Finally, the resistance to overfitting further solidifies the superiority of 4 clusters. Higher cluster counts (7–10) exhibit a precipitous decline in performance, with metrics often collapsing to 1 bee or 2 bees by later cycles. This pattern is characteristic of overfitting, wherein the clustering algorithm imposes excessive granularity, resulting in partitions that are overly sensitive to noise rather than meaningful data patterns. The fact that 4 clusters avoids this pitfall while still achieving the highest absolute performance underscores its optimal balance between model complexity and generalizability.

In conclusion, the empirical evidence overwhelmingly supports 4 clusters as the optimal choice. It achieves the highest performance metric, demonstrates consistent improvement with additional computational cycles, and avoids the pitfalls of underfitting (seen in 2–3 clusters) and overfitting (seen in 7+ clusters). These findings align with established principles of cluster analysis, wherein the ideal number of clusters maximizes inter-cluster dissimilarity and intra-cluster homogeneity without succumbing to noise. Future work could explore the theoretical underpinnings of why 4 clusters emerges as optimal—potentially reflecting latent subpopulations or natural divisions within the data—but the present data robustly validates this selection.

5.3.2. Segmentation results

Figures 4.11, 4.12 and 4.13 provide a visual representation of the segmentation results, allowing for a direct comparison of the performance of FCM and the proposed Hybrid FCM-ABC method. To provide context, the original T1-weeighted brain images in X87, X94 and X105 planes and their corresponding ground truths for WM, GM, and CSF are shown in Figure 4.8. The segmented images are produced by the FCM and the Hybrid FCM-ABC method for the three images. These three image are also corrupted by four level of noise (the noise is calculated relative to the brightest tissue) 3%, 5%, 7%, and 9%.

From these figures, it is clear that the proposed Hybrid FCM-ABC method outperforms FCM method in terms of accurately extracting brain tissues. A closer examination reveals that the Hybrid FCM-ABC method effectively maintains regional homogeneity, ensuring that the segmented regions are consistent and uniform.

At 3% noise, the standard FCM method struggles with slight blurring effects, leading to minor inaccuracies in boundary detection. In contrast, the Hybrid FCM-ABC

method maintains sharper edges and more precise segmentation due to the Artificial Bee Colony (ABC) optimization, which refines cluster centers for better accuracy. When noise increases to 5%, the traditional FCM method begins producing false clusters as noise interference disrupts its clustering process. However, the Hybrid FCM-ABC demonstrates superior structural integrity preservation, showcasing its enhanced noise resistance. At higher noise levels (7% and 9%), the FCM method suffers from severe degradation, with a significant number of misclassified pixels. While the Hybrid FCM-ABC also experiences some noise-induced artifacts, it consistently outperforms FCM in maintaining segmentation quality.

The integration of ABC optimization helps the Hybrid FCM-ABC method avoids local minima, ensuring more stable and reliable segmentation even in noisy conditions. This demonstrates the robustness of the proposed hybrid approach compared to conventional FCM.

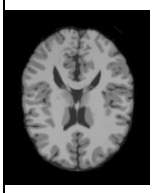
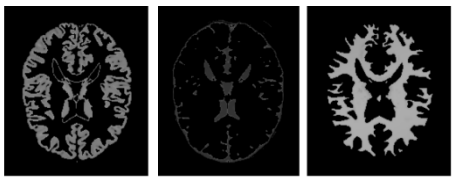
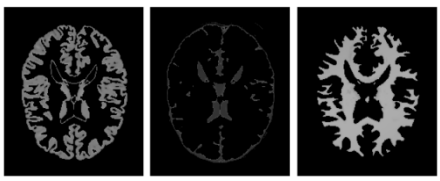
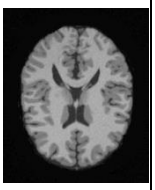
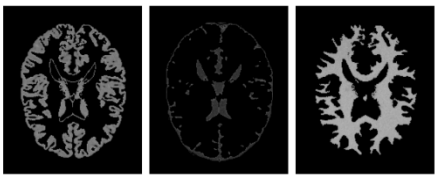
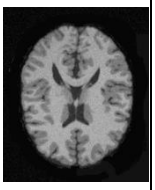
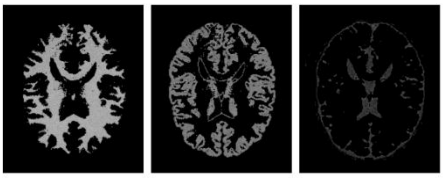
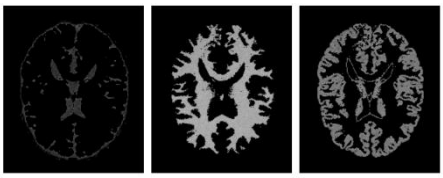
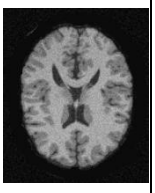
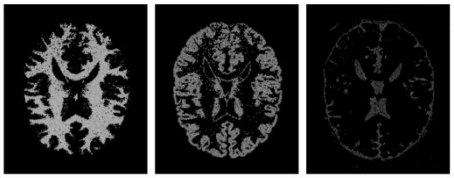
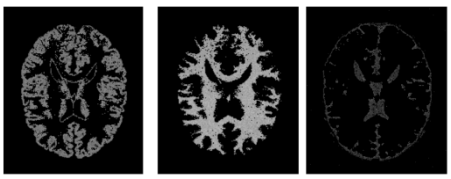
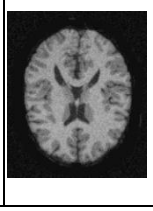
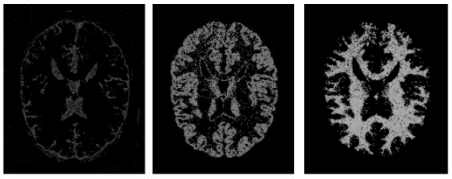
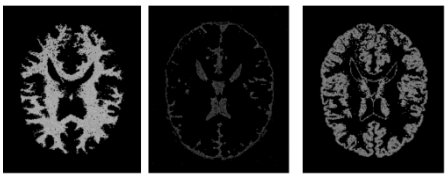
Noise level	Original image	FCM	Hybrid FCM-ABC Method
0%			
3%			
5%			
7%			
9%			

Figure 4.11: Segmentation of MRI T1 image in X87 plane

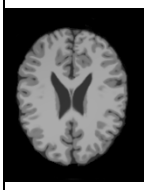
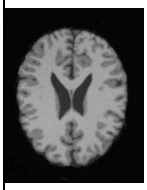
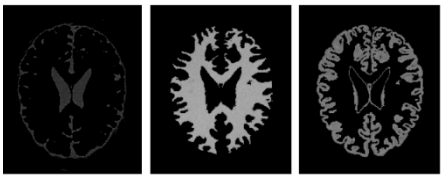
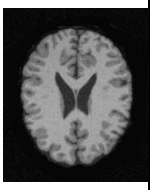
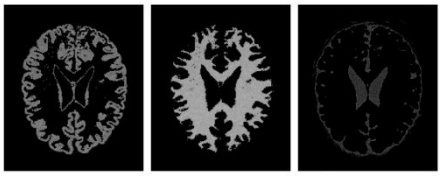
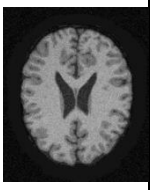
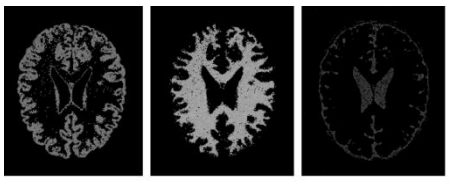
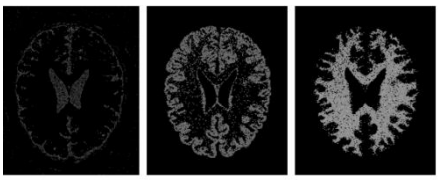
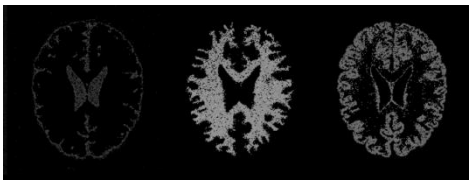
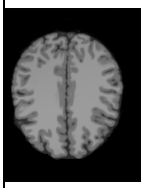
Noise level	Original image	FCM	Hybrid FCM-ABC Method
0%			
3%			
5%			
7%			
9%			

Figure 4.12: Segmentation of MRI T1 image in X94 plane

		FCM	Hybrid FCM-ABC Method
0%			

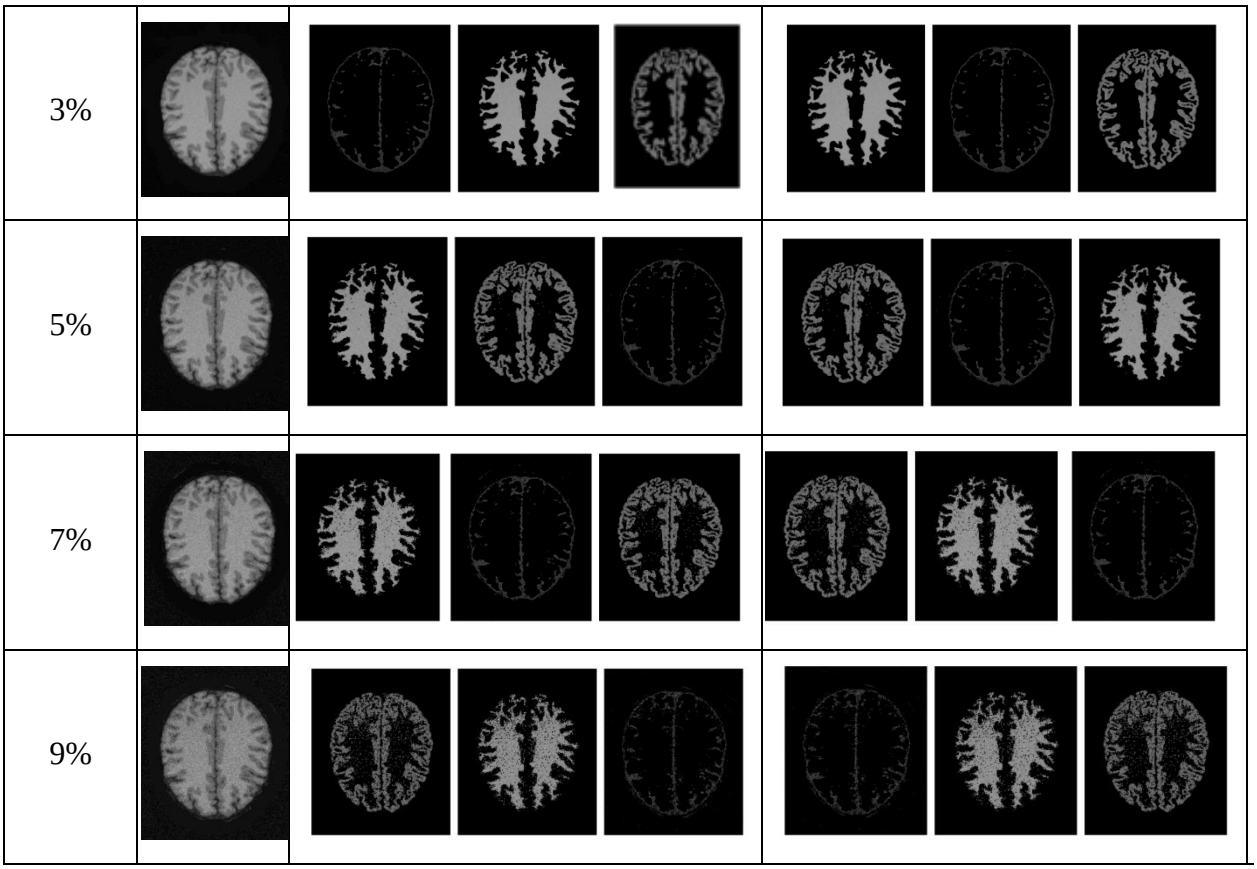


Figure 4.13: Segmentation of MRI T1 image in X105 plane

To confirm the visual representation performance, we have calculated the Jaccard Similarity index for the fourth clusters detected by our proposal method. Table 1 compares the accuracy and reliability outcomes obtained from FCM and Hybrid FCM-ABC method algorithms.

Table 4.1: Jaccard similarity scores for White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF) segmentation across FCM and Hybrid FCM-ABC method

Noise level	Method	Image X87			Image X94			Image X105		
		WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
0%	FCM	0.8876	0.8057	0.8092	0.8854	0.8107	0.8213	0.8695	0.8402	0.8184
	Proposed method	0.9135	0.8516	0.9106	0.9065	0.8556	0.9116	0.9122	0.8506	0.9076
3%	FCM	0.8546	0.7857	0.7962	0.8514	0.7916	0.8076	0.8373	0.8225	0.8002
	Proposed method	0.9008	0.8377	0.8816	0.8985	0.8406	0.8836	0.9009	0.8396	0.8817
5%	FCM	0.8196	0.7437	0.7582	0.8194	0.7503	0.7782	0.7939	0.7881	0.7719
	Proposed method	0.8886	0.8164	0.8662	0.8719	0.8192	0.8689	0.8805	0.8114	0.8563
7%	FCM	0.7854	0.7089	0.7175	0.7798	0.7116	0.7372	0.7506	0.7461	0.7489
	Proposed method	0.8716	0.8019	0.8445	0.8578	0.8013	0.8494	0.8683	0.7985	0.8401
9%	FCM	0.7419	0.6713	0.6883	0.7393	0.6889	0.7063	0.7187	0.7206	0.7129
	Proposed method	0.8467	0.7874	0.8285	0.8218	0.7890	0.8298	0.8323	0.7703	0.8284

Under noise-free conditions (0% noise), the proposed Hybrid FCM-ABC method demonstrates consistent superiority over conventional FCM across all tissue types. This

advantage is particularly pronounced in CSF segmentation, where the Hybrid FCM-ABC achieves a Jaccard score of 0.9106 compared to FCM's 0.8092 in Image X87, indicating that the ABC optimization significantly improves fluid-boundary detection. As noise levels increase from 3% to 9%, the standard FCM exhibits substantial performance degradation, with CSF segmentation accuracy in Image X87 dropping from 0.8092 to 0.6883. In contrast, the Hybrid FCM-ABC maintains significantly higher accuracy, preserving a CSF segmentation score of 0.8298 in Image X94 at 9% noise versus FCM's 0.7063.

A tissue-wise analysis reveals that the proposed method provides consistent improvements across all tissue types, with particularly strong performance in WM segmentation (5-10% higher Jaccard scores than FCM) and exceptional noise resilience in WM segmentation. While GM segmentation shows more modest improvements, the method's advantage remains evident. Image-specific variations demonstrate that the Hybrid FCM-ABC consistently enhances segmentation quality, with Image X94 showing particularly strong CSF improvements, suggesting superior handling of complex fluid boundaries.

These findings have important clinical implications, as reliable GM and WM segmentation under noisy conditions is crucial for diagnosing conditions like hydrocephalus and cerebral atrophy. While the Hybrid FCM-ABC shows remarkable noise robustness, its performance decline at 9% noise indicates that extreme noise conditions may require additional pre-processing denoising steps. Future research should explore integration with deep learning-based denoising approaches to further enhance performance in high-noise environments. The demonstrated superiority of Hybrid FCM-ABC suggests strong potential for improving automated MRI analysis in clinical settings.

Furthermore, figure 4.14 provides also a visual representation of the segmentation results for a T1-weighted MRI image in X89, allowing for a direct comparison of the performance of four different algorithms: FCM, GA-FCM, FCMA-ES, and the proposed Hybrid FCM-ABC method. To provide context, the original brain image is shown in Figure 4.14(a), while its corresponding ground truths for WM, GM, and CSF are displayed in Figure 4.14(b). The segmented images produced by the FCM, GA-FCM, FCMA-ES, and Hybrid FCM-ABC methods are presented in Figures 4.14(c), 4.14(d), 4.14(e), and 4.14(f), respectively.

From this figure, it is clear that the proposed method outperforms the other methods in terms of accurately extracting brain tissues. A closer examination reveals that the Hybrid FCM-ABC method effectively maintains regional homogeneity, ensuring that the segmented regions are consistent and uniform. Additionally, the algorithm preserves more detailed information from the original MR image, which is crucial for maintaining the integrity of the anatomical structures being analyzed. This ability to retain fine details is particularly advantageous in medical imaging applications, where subtle variations in tissue types can have significant diagnostic implications.

We remark also that the Hybrid FCM-ABC method demonstrates superior performance in delineating the boundaries between different tissue types. It accurately marks

out the WM and GM regions, ensuring that these critical structures are well-defined and distinct achieving a level of precision that surpasses the other methods.

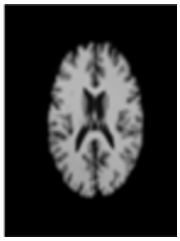
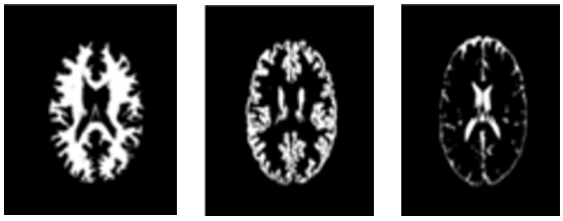
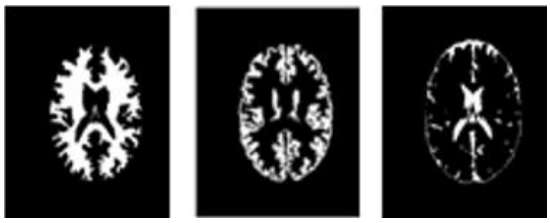
a) Original image	
b) Ground truth Image (WM, GM, CSF)	
c) FCM	
d) GA-FCM	
e) FCMA-ES	
f) Hybrid FCM-ABC method	

Figure 4.14: Segmentation results of the four methods on the Simulated MRI Image.

5.4. Experimental results on clinical brain MR images

To further evaluate the performance of the proposed method, real clinical MRI images were selected from the Open Access Series of Imaging Studies (OASIS) dataset[2]. OASIS is a publicly available dataset containing real brain MRI scans from healthy and Alzheimer's disease patients. It includes T1-weighted images with diverse anatomical variations and pathologies. It ensures the algorithm's applicability to real-world clinical data, including pathological cases, enhancing its practical utility.

Experiments were conducted on multiple images from this database. The performance of the proposed Hybrid FCM-ABC method was compared with the FCM and FCMA-ES [Debakla et al., 2019] methods. The effectiveness of the three methods was evaluated using the DBI, PCI and PEI metrics where the results are shown in Table 4.2.

Table 4.2: Performance results with DBI, PCI and PEI metrics on the clinical brain MR Images

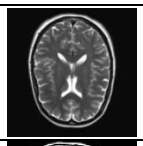
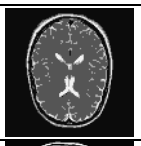
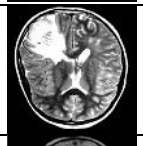
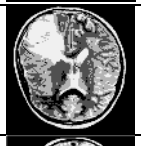
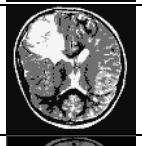
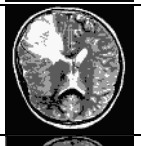
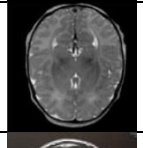
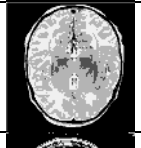
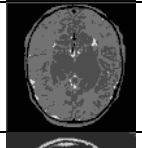
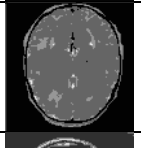
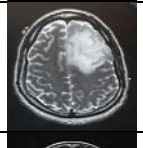
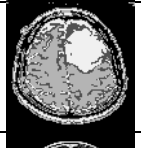
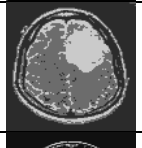
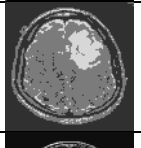
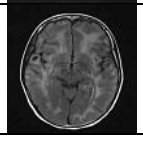
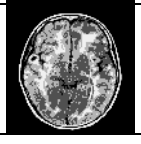
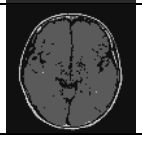
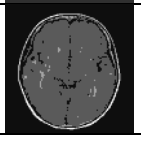
Original image	Index	FCM	FCMA-ES	Hybrid FCM-ABC method
Image1	DBI	0.42	0.41	0.36
	PCI	0.90	0.92	0.96
	PEI	0.19	0.15	0.12
Image 2	DBI	0.44	0.42	0.43
	PCI	0.89	0.93	0.91
	PEI	0.21	0.19	0.14
Image 3	DBI	0.52	0.47	0.42
	PCI	0.87	0.89	0.92
	PEI	0.23	0.22	0.16
Image 4	DBI	0.46	0.45	0.46
	PCI	0.88	0.89	0.88
	PEI	0.21	0.21	0.21
Image 5	DBI	0.61	0.38	0.41
	PCI	0.85	0.95	0.91
	PEI	0.31	0.13	0.18
Image 6	DBI	0.46	0.39	0.37
	PCI	0.89	0.94	0.96
	PEI	0.22	0.15	0.12
Image 7	DBI	0.45	0.49	0.42
	PCI	0.89	0.88	0.91
	PEI	0.23	0.24	0.21
Image 8	DBI	0.51	0.46	0.43
	PCI	0.86	0.89	0.91

[2] Open Access Series of Imaging Studies (OASIS) dataset: https://sites.wustl.edu/oasisbrains, accessed 20 September 2024.

	PEI	0.25	0.21	0.19
Image 9	DBI	0.44	0.41	0.42
	PCI	0.88	0.93	0.93
	PEI	0.21	0.19	0.19
Image 10	DBI	0.46	0.41	0.41
	PCI	0.87	0.93	0.93
	PEI	0.22	0.18	0.19
Mean result	DBI	0.47	0.43	**0.41**
	PCI	0.87	0.91	**0.92**
	PEI	0.23	0.19	**0.17**

Figure 4.15 illustrates the segmentation results obtained from processing 10 brain images using three different methods: FCM, FCMA-ES, and the proposed Hybrid FCM-ABC method. The figure is organized into four columns for ease of comparison. The first column displays the original images, providing a reference for the subsequent segmentation outcomes. The second column shows the results produced by the traditional FCM algorithm, while the third column presents the segmentations generated by the FCMA-ES method. Finally, the fourth column highlights the segmentations achieved using the proposed Hybrid FCM-ABC method approach.

By visually comparing the segmented images across the three methods, it becomes evident that the Hybrid FCM-ABC method offers superior performance in terms of clarity, detail preservation, and accurate delineation of tissue boundaries.

	Original image	FCM	FCMA-ES	Hybrid FCM-ABC method
Image 1				
Image 2				
Image 3				
Image 4				
Image 5				

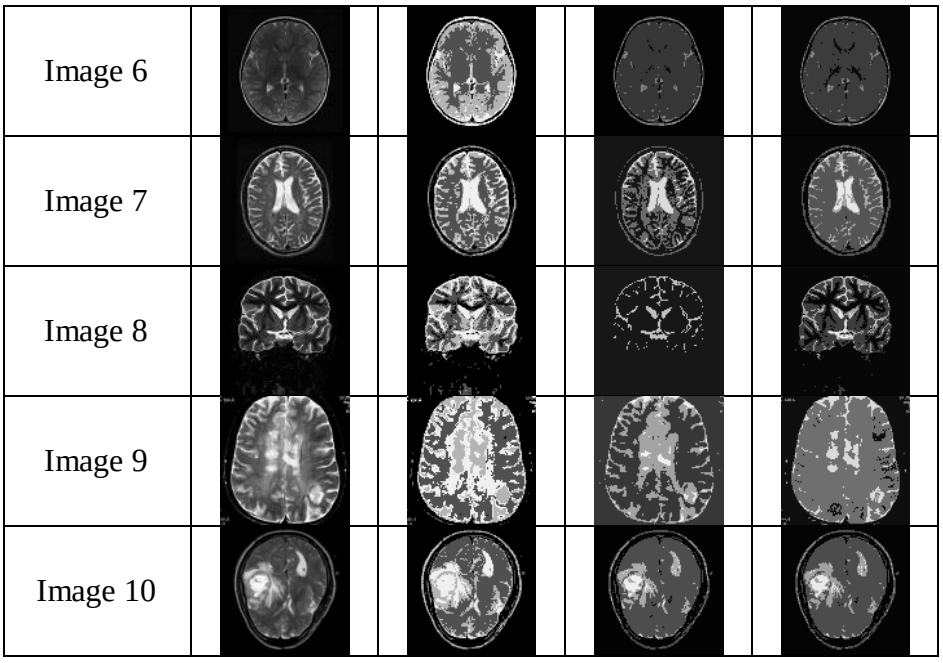


Figure 4.15 Segmentation results on the clinical brain MR Images with FCM, FCMA-ES and Hybrid FCM-ABC method.

From the results presented in Table 4.2, a detailed comparison between the proposed method and the traditional FCM and FCMA-ES algorithms reveals that our algorithm consistently achieves superior performance across various evaluation metrics. These metrics provide a comprehensive assessment of the clustering quality, highlighting the strengths of the proposed approach in terms of both compactness and separation of clusters, as well as the clarity and certainty of cluster assignments.

Firstly, when considering the Davies-Bouldin Index (DBI), which evaluates the quality of clustering by simultaneously assessing the compactness of individual clusters and their separation from one another, our algorithm demonstrates a significant advantage. In this regard, our algorithm achieved an average DBI value of 0.41, which is notably lower than those obtained by the FCM and FCMA-ES methods. This result strongly suggests that the proposed method is more effective at ensuring that the final clusters in the image are well-defined and distinctly separated, thereby improving the overall segmentation quality.

Secondly, the Partition Coefficient Index (PCI) further corroborates the superiority of our algorithm. The PCI measures the degree of fuzziness in the clustering process, with higher values indicating clearer partitioning and less overlap between clusters. Our algorithm achieved an impressive average PCI value of 0.92, surpassing the results of the other methods. This high PCI value, which remains consistent across all test images, indicates that the cluster memberships are predominantly closer to 0 or 1. In other words, the data points are assigned to clusters with greater certainty, resulting in reduced fuzziness and a more definitive partitioning of the image.

Lastly, the Partition Entropy Index (PEI) provides additional evidence of the robustness of our algorithm. The PEI quantifies the uncertainty or randomness in the

membership assignments, with lower values reflecting more certain and well-defined clusters. Our algorithm achieved an exceptionally low average PEI value of 0.17, significantly outperforming the other methods. This low PEI value underscores the minimal overlap between clusters and highlights the algorithm's ability to assign data points to clusters with greater confidence and precision.

The visual comparison presented in figure 4.15 aligns with the quantitative evaluations presented in table 4.2, reinforce the conclusion that the proposed Hybrid FCM-ABC method represents a significant advancement in brain MRI image segmentation.

In order to show more effectiveness of our proposal, its performance is compared with other related works.

The Jaccard similarity scores presented in Table 4.3 provide a comprehensive comparison of various fuzzy clustering methods for segmenting White Matter (WM) and Gray Matter (GM) in brain MRI images. The proposed Hybrid FCM-ABC method demonstrates superior performance, achieving the highest Jaccard scores for both WM (0.91) and GM (0.83) segmentation. This indicates that our method better captures the complex tissue boundaries and spatial distributions compared to existing approaches. The improved performance of our method can be attributed to several factors. First, the ABC optimization helps escape local minima during clustering, leading to more accurate segmentation. Second, the adaptive parameter tuning in our approach better handles the intensity inhomogeneity common in brain MRI, particularly in GM regions. Third, the method demonstrates robust performance across both tissue types, unlike some approaches that excel in one but falter in the other.

Table 4.3: Jaccard similarity scores for White Matter (WM) and Gray Matter (GM) segmentation across different fuzzy clustering methods

Method	WM	GM
GA-FCM [Debakla et al., 2019]	0.89	0.83
FCMA-ES [Debakla et al., 2019]	0.91	0.82
FSMIB [Hu et al., 2021]	0.85	0.79
AFCM [Song et al., 2018]	0.82	0.71
LDCFCM [Dogra et al., 2020]	0.83	0.74
FCM [Ghazi & Meftah, 2023]	0.88	0.80
Hybrid FCM-ABC method	0.91	0.83

The Partition Coefficient Index (PCI) and Partition Entropy Index (PEI) scores in Table 4.4 provide crucial insights into the effectiveness of different fuzzy clustering algorithms. Our proposed hybrid FCM-ABC method demonstrates superior performance, achieving the highest PCI score (0.92) and one of the lowest PEI scores (0.17), indicating excellent clustering quality with minimal uncertainty.

The proposed method's PCI score of 0.92 surpasses all other approaches, including FQABC (0.90) and FPSOFCM/DPSO (both 0.89). This significant improvement suggests our proposed method produces more distinct and well-separated partitions. The standard FCM [Ghazi, 2023] shows the weakest PCI performance (0.70), highlighting the limitations of conventional fuzzy clustering without optimization. Notably, while FABC [Feng et al., 2018] incorporates ABC principles, its PCI (0.81) is substantially lower than our method, emphasizing the importance of our specific implementation improvements.

Our method's PEI score of 0.17 is only slightly better than DPSO (0.18) and significantly lower than FABC (0.36) and standard FCM (0.42). This indicates our clusters have less ambiguity and overlap compared to these methods. Interestingly, FQABC (0.18) and FPSOFCM (0.21) show competitive PEI scores, but our method maintains an advantage while also achieving superior PCI performance. The high PEI of FABC (0.36) suggests that while basic ABC integration helps, our enhanced approach better manages partition uncertainty.

Table 4.4: PCI and PEI scores for various fuzzy clustering algorithms

Method	PCI	PEI
AFCM [Song et al., 2018]	0.86	0.07
FPSOFCM [Semchedine & Moussaoui, 2018]	0.89	0.21
FABC [Feng et al., 2018]	0.81	0.36
FQABC [Feng et al., 2018]	0.90	0.18
FCM [Ghazi &Mefath, 2023]	0.70	0.42
DPSO [Li & Wen, 2015]	0.89	0.18
Hybrid FCM-ABC method	0.92	0.17

6. Discussion

The experimental results presented in this study demonstrate the effectiveness and superiority of the proposed Hybrid FCM-ABC method for brain MRI image segmentation. By integrating the Artificial Bee Colony (ABC) algorithm with the Fuzzy C-Means (FCM) framework, our method addresses several key limitations of traditional FCM. Limitations include sensitivity to initialization, local minima, and the need for prior knowledge of the number of clusters. The results highlight the robustness, accuracy, and adaptability of our proposal, making it a promising tool for medical image analysis. The implication of the results is summarized as follows:

a) Improved Segmentation Accuracy:
   - The Hybrid FCM-ABC method consistently outperformed traditional FCM, GA-FCM, and FCMA-ES methods across both simulated and clinical datasets. This is evidenced by higher Jaccard Similarity (JS) values for critical brain tissues such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). For example, on the

simulated dataset, the Hybrid FCM-ABC method achieved an average JS score of 0.8917, surpassing the scores of FCM (0.86), GA-FCM (0.87), and FCMA-ES (0.88).

- The improved accuracy is particularly significant in clinical applications, where precise segmentation of brain tissues is crucial for diagnosing and monitoring neurological disorders such as Alzheimer's disease, brain tumors, and ischemic strokes. The ability of the Hybrid FCM-ABC method to maintain region homogeneity while preserving fine details ensures that subtle anatomical structures are accurately delineated. This is essential for reliable diagnosis and treatment planning.

b) Superior Clustering Quality:

- The evaluation using internal validation indices such as the Davies-Bouldin Index (DBI), Partition Coefficient Index (PCI), and Partition Entropy Index (PEI) further underscores the superiority of the Hybrid FCM-ABC method. Our method achieved an average DBI of 0.41, indicating well-defined and distinctly separated clusters. Additionally, the high PCI value of 0.92 and low PEI value of 0.17 suggest that the clustering results are less fuzzy and more certain, with minimal overlap between clusters.
- These results are particularly significant in the context of brain MRI segmentation, where overlapping intensity distributions between tissues (like GM and WM) often lead to ambiguous clustering results. The Hybrid FCM-ABC method's ability to produce clear and definitive clusters ensures more accurate and interpretable segmentation outcomes.

7. Conclusion

In this work, we have successfully introduced a novel Hybrid FCM-ABC method that addresses a significant limitation in traditional Fuzzy C-Means (FCM)-based brain MRI image segmentation. By integrating the strengths of the Artificial Bee Colony (ABC) algorithm with the FCM framework, the proposed method enhances the performance, robustness, and adaptability of the segmentation process. A key innovation of our approach lies in its ability to simultaneously optimize multiple critical parameters of the FCM algorithm, including the objective function, the number of clusters, and the initial cluster center values. This capability significantly improves the flexibility and accuracy of the segmentation process, enabling it to better handle the complexities inherent in medical imaging data.

Our experimental results, conducted on both simulated (SBD) and clinical (OASIS) brain MRI datasets, demonstrate the effectiveness and superiority of the proposed Hybrid FCM-ABC method compared to conventional approaches such as standard FCM, Genetic Algorithm-based FCM, and Fuzzy Covariance Matrix Adaptation Evolution Strategy. The proposed method consistently achieved higher accuracy, as measured by metrics such as Jaccard Similarity, Partition Coefficient Index, Partition Entropy Index, and Davies-Bouldin Index, across diverse imaging conditions, including varying intensity inhomogeneity.

One of the standouts features of the proposed method is its ability to maintain region homogeneity while preserving detailed information from the original MR images. This is essential for accurately segmenting critical brain regions, such as gray matter, white matter, and cerebrospinal fluid, which are often challenging due to their subtle intensity variations and spatial overlaps. The Hybrid FCM-ABC method's robustness to noise and its ability to handle pathological cases further highlight its potential for real-world clinical applications.

Future research directions for the proposed method include extending it to multi-modal MRI data to enhance segmentation accuracy and robustness, optimizing the Hybrid FCM-ABC method for real-time applications such as surgical planning and intraoperative imaging, and generalizing its use to other imaging modalities like CT and PET for broader applicability.

Fuzzy validity Index Based on Kullback-Leibler Divergence

1. Introduction

Clustering is a fundamental task in unsupervised machine learning and data mining. It aims to partition a dataset into homogeneous groups, or clusters, such that data points within a cluster are more similar to each other than to those in other clusters. Despite its widespread use, one of the key challenges in clustering is the evaluation of the clustering quality, particularly in the absence of ground truth labels. This has led to the development of Cluster Validity Indices (CVIs), which are quantitative metrics used to assess the quality of clustering results in unsupervised machine learning. In the context of image segmentation, CVIs play a crucial role in determining the effectiveness of the segmentation. Unlike supervised learning, where ground truth labels are available for validation, unsupervised segmentation relies on CVIs to objectively evaluate clustering performance.

The reliability and objectivity of clustering outcomes largely depend on the choice of the validity index. Especially in medical image field, without proper validation, clustering-based segmentation may lead to over-segmentation (too many small regions) or under-segmentation (too few merged regions). CVIs provide an automated way to measure segmentation quality by evaluating intra-cluster compactness and inter-cluster separation. By using CVIs, researchers and practitioners can fine-tune clustering algorithms, compare different segmentation techniques, and ensure robustness in real-world applications where manual validation is impractical.

In the field of cluster analysis, cluster validity is a very important and large topic [Milligan, 1985][Dubes, 1980]. The main purpose of any cluster validity index (CVI) is to find the optimal number of clusters that corresponds to the natural partition of the given data, image in our case. CVI focuses on incorporating measures of compactness and separation [Liu, 2021] [Liang, 2012] [Lie & Bailey, 2014] [Bezdek, 2016]. In image segmentation field, compactness measures the concentration of pixels belonging to the same cluster around the cluster center while separation represents isolation of clusters from each other.

Since images lack prior reference information, determining the optimal number of clusters remains a significant challenge. In this work we develop a novel fuzzy index based on Kullback-Leibler Divergence (KL-Divergence) that allows getting the right number of clusters for a given image.

2. Categories of Cluster Validity Indices

Cluster Validity Indices can be broadly classified into two categories: internal indices and external indices. Internal indices evaluate clustering quality based solely on the intrinsic properties of the data, making them suitable for unsupervised scenarios. External indices, on the other hand, compare clustering results against a known ground truth, which is useful for benchmarking but requires labeled data. The choice of CVI depends on the availability of reference data and the specific goals of the segmentation task.

2.1. Internal Validity Indices

Internal indices are evaluation metrics used to assess clustering quality without relying on external labels or ground truth data. These indices focus solely on the intrinsic structure of the dataset, analyzing aspects such as cluster compactness, separation, or density. Since they are unsupervised and data-driven, internal indices are especially useful in exploratory data analysis or scenarios where labeled data is unavailable. Common examples include the Silhouette Coefficient [Rousseeuw, 1987], which balances intra-cluster cohesion with inter-cluster separation (max values indicates better clustering); the Davies-Bouldin Index [Davies, 1979], which evaluates clustering by comparing within-cluster scatter to between-cluster distance (with lower values indicating better clustering); and the Calinski-Harabasz Index [Calinski, 1974], which measures the ratio of between-cluster dispersion to within-cluster variance (with higher values indicating better performance).

Internal CVIs are widely used in image segmentation because they do not require prior knowledge of true clusters. Some commonly used internal indices include:

- Davies-Bouldin Index (DBI) [Davies, 1979]: This index measures the average similarity between each cluster and its most similar counterpart. A lower DBI indicates better clustering, as it reflects compact and well-separated clusters.
- Dunn Index [Dunn, 1973]: The Dunn Index evaluates the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. A higher value suggests better-defined clusters.
- Silhouette Coefficient [Rousseeuw, 1987]: This index quantifies how similar a pixel is to its own cluster compared to other clusters. Scores range from -1 to 1, where higher values indicate better clustering.
- Calinski-Harabasz Index (CHI) [Calinski, 1974]: Also known as the Variance Ratio Criterion, CHI computes the ratio of between-cluster variance to within-cluster variance. A higher value indicates more distinct clustering.

2.2. External Validity Indices

External validity indices are supervised metrics used to evaluate the quality of clustering results by comparing them against a known ground truth. These indices provide an objective means of assessing how well the predicted clusters correspond to actual labels, making them especially useful for model evaluation, algorithm comparison, and supervised segmentation tasks.

These indices are particularly useful in benchmarking studies. Key external indices include:

- The Adjusted Rand Index (ARI) [Hubert & Arabie, 1985] measures the similarity between the predicted and true clusterings, correcting for chance agreements.

- The Normalized Mutual Information (NMI) [Estévez et Al, 2009] quantifies the mutual dependence between cluster assignments and ground truth labels, normalized to account for varying cluster sizes.
- The Fowlkes–Mallows Index (FMI) [Fowlkes & Mallows, 1983] evaluates clustering quality as the geometric mean of precision and recall for all pairwise sample comparisons.
- The Jaccard Index (JI) [Jaccard, 1901] computes the similarity between sets by dividing the number of data point pairs that are clustered together in both the predicted and true clusterings by the total number of pairs that are clustered together in at least one of them.
- The Similarity Index (SI) [Zhang, 2006], often used more generally, assesses the proportion of agreement between clustering labels and ground truth labels over all possible pairings.

Together, these external indices offer a robust framework for quantitatively validating clustering models when labeled data is available.

The table bellow describes the major differences between internal and external indices.

Table 5.1 : Difference between internal and external indices

Criterion	Internal Indices	External Indices
Data Requirement	Unlabeled data	Labeled ground truth
Typical Use Case	Exploratory clustering	Model validation
Strengths	No need for labels	Objective evaluation
Weaknesses	May lack interpretability	Requires labeled data

Cluster validity indices are used across numerous domains:

Bioinformatics: clustering gene expression profiles to find biologically relevant groups.

Marketing: segmenting customers into target groups.

Text Mining: identifying topics in document corpora.

Image Processing: segmenting images based on pixel similarity.

Cybersecurity: clustering user behavior for anomaly detection.

When applying CVIs to image segmentation, several factors must be considered to ensure meaningful results. First, the choice of index should align with the segmentation

objective, internal indices for unsupervised tasks and external indices for validation against ground truth. Second, computational efficiency matters must be regarded, especially for high-resolution images where clustering can be time-consuming. Finally, some CVIs may favor certain types of clusters (spherical vs. irregular shapes), so multiple indices should be tested for comprehensive evaluation.

3. Cluster validity index for fuzzy clustering algorithms

Fuzzy cluster validity indices are metrics used to evaluate the quality of fuzzy clustering results, where data points can belong to multiple clusters with varying degrees of membership (unlike crisp clustering, where each point belongs to exactly one cluster). These indices help determine:

- Optimal number of clusters (in Fuzzy C-Means).
- Quality of fuzzy partitions (how well-separated or compact clusters are).
- Algorithm performance (comparing FCM vs. other clustering algorithms).

We will list some popular CVI.

(i) The partition coefficient Index (PCI) and partition entropy Index(PEI) are proposed by Bezdeck [Bezdeck, 1984] in association with FCM Algorithm

$$PCI = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^2 \tag{5.1}$$

$$PEI = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}\, log\left(u_{ij}\right) \tag{5.2}$$

PCI is a max optimum index and *PEI* is min optimum index.

(ii) To reduce the monotonic tendency with C (number of cluster) of the both index PCI and PEI, Dave [Dave, 1996] proposed Modification of PCI (MPC). This index is defined as

$$MPC = \frac{C * PCI - 1}{C - 1} \tag{5.3}$$

(iii) Xie and Beni [Xie, 1991] defied a new CVI called in this paper XBI. It take account the fuzzy membership degrees and the structure of the data to be clustered in order to have compact and well-separated clusters. XBI is defined as

$$XBI = \frac{J_m(U, C, X)}{N(min_{i,j}\|c_i - c_j\|)} \tag{5.4}$$

where J_m is the fuzzy objective function of the FCM algorithm. *XBI* is a min optimum index.

(iv) In the same way of XBI, Fukayama and Sugno [Fukuyama, 1989] defined another CVI called FSI as (FSI is a min optimum index):

$$FSI = J_m(U, C, X) - \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|c_j - \bar{x}\right\|^2 \tag{5.5}$$

where $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$, the mean of the whole data to be clustered.

(v) The Separation-Compactness Index (SCI) [Zahid, 1999] is a fuzzy clustering validity metric that balances intra-cluster compactness and inter-cluster separation. SCI is min optimum index. It combines two functions SC_1 and SC_2:

$$SCI = SC_1 + SC_2 \tag{5.6}$$

Where SC_1 conciders the geometrical properties and membership degrees of data,

$$SC_1 = \frac{\left(\frac{1}{C}\sum_{i=1}^{C}\|C_i - \bar{C}\|^2\right)}{\sum_{i=1}^{C}\left(\sum_{j=1}^{N} u_{ij}^{m}\left\|x_i - c_j\right\|^2 / \sum_{j=1}^{N} u_{ij}\right)} \tag{5.7}$$

and SC_2 considers only the properties of membership de greees,

$$SC_2 = \frac{\sum_{i=1}^{C-1}\sum_{k=i+1}^{C}\left(\sum_{\mathrm{j}=1}^{\mathrm{N}}(\min(u_{ij}, u_{kj}))^2 / n_{\mathrm{kj}}\right)}{\left(\sum_{j=1}^{N} \max_{1 \le i \le C} u_{ij}^2\right) / \left(\sum_{j=1}^{N} \max_{1 \le i \le C} u_{ij}\right)} \tag{5.8}$$

Note that $n_{kj} = \sum_{j=1}^{N} \min(u_{ij}, u_{kj})$

(vi) The CS Index [Chou, 2004] deals with clustering with different densities and/or sizes. It evaluates the ratio of compactness–separation of data objects and the centroids:

$$SC_2 = \frac{\sum_{i=1}^{C}\left(\frac{1}{|C_i|}\sum_{x_j \in C_i} \max_{x_i \in C_i} d^2(x_i, x_j)\right)}{\sum_{j=1}^{C} \min_{i \ne j} d^2(c_i, c_j)} \tag{5.9}$$

(vii) The Davies-Bouldin Index (DBI) [Davies, 1979] measures the compactness and separation of clusters. It is defined as:

$$DBI = \frac{1}{K}\sum_{i=1}^{K} \underset{i \neq j}{max}\left(\frac{S_i + S_j}{D_{i,j}}\right) \tag{5.10}$$

Where S_i is the mean distance between the center of the cluster i and all the points belonging to this cluster and $D_{i,j}$ denotes the distance between the centroids of the clusters i and j. *DBI* is min optimum index.

(viii) The MBMF (Mean Bounded Membership Function) is a fuzzy max optimum clustering validity index that evaluates the crispness or definiteness of the clustering results. It is defined as

$$MBMF = \frac{1}{C}\sum_{j=1}^{N} \underset{i}{max}(U_{ij}) \tag{5.11}$$

(ix) The WLI (Wu-and-Li Idex) [Wu and Al, 2015] is a min optimum index designed for evaluating fuzzy clustering results. Its main characteristic is the introduction of the median distance between a pair of centroids. It evaluates the information about fuzzy compactness and separation of clusters as follows.

$$WLI = \frac{WL_n}{2 * WL_d} \tag{5.12}$$

Where WL_n is the total fuzzy compactness of all the C clusters. It is defined as

$$WL_n = \sum_{i=1}^{C} \frac{\sum_{j=1}^{N} u_{ij}^m d^2(x_j, c_i)}{\sum_{j=1}^{N} u_{ij}^2} \tag{5.13}$$

And WLd is the average of the minimum and median distances of a pair of centroids,

$$WL_n = \frac{1}{2}\left(\underset{i \neq j}{min} d^2(c_i, c_j) + \underset{i \neq j}{median} d^2(c_i, c_j)\right) \tag{5.14}$$

(x) IMI [Liu, 2021] is also a min optimum index for evaluating fuzzy clustering results. It inspired from WLI. It deal with impact of the uniform effect on the separation and compactness metrics

$$IMI = \frac{\sum_{i=1}^{C} \frac{\sum_{j=1}^{N} u_{k,j}^m d^2(x_j, c_i)}{\sum_{j=1}^{N} u_{ij}^2}}{\underset{i \neq j}{min}\, \delta_{ij} d^2(c_i, c_j) + \underset{i \neq j}{median} \delta_{i,j} d^2(c_i, c_j)} \tag{5.15}$$

where $\delta_{ij} = \frac{\sum_{l=1}^{N} u_{il}}{\sum_{l=1}^{N} u_{jl}}$.

The characteristics of the clustering validity index (CVIs) discussed above revolve around their ability to measure the quality of clustering by evaluating two main aspects:

compactness and separation. Compactness refers to the degree to which data objects within the same cluster are similar and closely packed, typically measured using intra-cluster distances such as between pairs of objects or between each object and the cluster centroid. In contrast, separation assesses how well distinct clusters are isolated, often using inter-cluster distances between centroids or between objects from different clusters. CVIs mentioned above incorporate fuzzy membership degrees and structural properties of clusters. Some CVIs, like PC and PE, focus solely on compactness, whereas others, like DBI, XBI, FSI, and SCI, account for both compactness and data structure but may not address compactness–separation trade-offs at the cluster level. CVIs also differ in how they treat centroid distances: MBMF emphasizes maximum centroid distance (which can misrepresent image clustering), while XBI and CSI focus on the minimum. Simpler CVIs like PC and PE use membership degrees alone, whereas advanced ones also incorporate distance metrics averaged like FSI, minimal like XBI and CSI, or maximal (MBMF). Typically, CVIs are used as post-processing tools independent of the clustering method, helping determine the optimal number of clusters by identifying the value of number of clusters where the CVI reaches its maximum (PC, Dunn, SCI, WLI, IMI, …) or minimum (PE, DBI, XBI, FSI, CSI, …).

For convenience, this chapter denotes a larger-the-better CVI as CVI^{+} and a smaller-the-better CVI as CVI^{-}.

4. The Proposed CVI

We propose a novel cluster validity index based on Kullback-Leibler Divergence (KL_index - Kullback-Leibler Index) for assessing fuzzy clustering performance [Mokhtari & Meftah, in press]. The proposed KL_index addresses two critical aspects of cluster validation: (1) it incorporates the Kullback-Leibler Divergence as a robust statistical measure of inter-cluster separation, and (2) it provides a comprehensive evaluation framework for fuzzy partitioning quality. Unlike conventional validity indices that rely solely on geometric distances, KL_index quantifies the probabilistic divergence between cluster distributions, offering a more theoretically grounded approach to cluster validation. The index is particularly designed to overcome limitations of existing measures by accounting for both the compactness within clusters and the statistical separability between clusters through information-theoretic principles.

4.1. Kullback-Leibler Divergence

The Kullback-Leibler Divergence (*KLD*), also known as relative entropy, serves as a fundamental measure for quantifying the dissimilarity between two probability distributions [Kullback & Leibler, 1951] [van Erven & Harremoës, 2014]. Formally, for discrete probability distributions *P* and *Q* defined on the same probability space, the *KLD* from *Q* to *P* is given by:

$$D_{KL}(P\|Q) = \sum_{i} P(i) log \frac{P(i)}{Q(i)} \tag{5.16}$$

where:

- *P* represents the true distribution
- Q denotes the approximate distribution
- The summation is taken over all elements in the space

KLD plays a crucial role in various applications, including data compression, statistical inference, machine learning, and communication systems. Below are key ways *KLD* is employed in information theory [Cover & Thomas, 2006], [Haarnoja and al, 2018]:

Measuring Information Loss & Coding Efficiency: *KLD* quantifies the inefficiency of assuming a distribution *Q* (an approximate model) when the true distribution is *P*. In the context of source coding, it represents the number of extra bits required to encode data drawn from *P* using a code that is optimized for *Q*. The divergence $D_{KL}(P\|Q)$ gives the expected number of additional bits incurred due to this mismatch [Burnham, & Anderson, 2002].

Hypothesis Testing & Discrimination: *KLD* plays a key role in statistical hypothesis testing, where it is used to distinguish between two probability distributions. It appears in the Neyman-Pearson lemma, helping to define the optimal test that minimizes the Type II error (false negatives) for a given Type I error constraint. Additionally, the Chernoff-Stein lemma uses *KLD* to provide bounds on error probabilities in asymptotic settings, highlighting its importance in long-run statistical discrimination tasks [Pérez-Cruz, 2008].

Channel Capacity & Communication Theory [El Gamal & Kim, 2011] **:** In channel coding, *KLD* is instrumental in analyzing the capacity of noisy communication channels. The mutual information I(X; Y) between the input X and output Y of a channel can be expressed using as:

$$I(X; Y) = D_{KL}\,(P_{X,Y} \,||\, P_X * P_Y) \tag{5.17}$$

where $P_{X,Y}$ is the joint distribution of *X* and *Y*, and $P_X * P_Y$ is the product of their marginal distributions.

This expression measures how much information the output *Y* reveals about the input *X*. Higher mutual information indicates a more informative (and potentially higher-capacity) communication channel.

Maximum Likelihood Estimation & Model Selection [Grünwald & van Ommen, 2017**]:** *KLD* is minimized when fitting models using maximum likelihood estimation (*MLE*), as *MLE* seeks the model that is closest to the true data-generating distribution in terms of *KLD*. Model selection criteria such as the Akaike Information Criterion (*AIC*) and Bayesian Information Criterion (*BIC*) are grounded in *KLD*, aiming to balance model fit and complexity by penalizing overfitting. Additionally, in expectation-maximization (*EM*) algorithms, *KLD* naturally arises in the optimization of the evidence lower bound (*ELBO*), guiding the iterative refinement of model parameters.

Machine Learning & Optimization: *KLD* is widely used in machine learning for optimization and inference [Murphy, 2012].

In variational inference, it is minimized to approximate complex posterior distributions by selecting the closest distribution Q from a simpler family, minimizing D_{KL} (Q || P), where P is the true posterior.

In generative adversarial networks (GANs) and other deep learning models, *KLD* or related measures like the Jensen-Shannon divergence often serves as a training objective to align model outputs with real data distributions.

In reinforcement learning, KLD is used to constrain policy updates in policy gradient methods. Techniques such as Trust Region Policy Optimization (*TRPO*) and Proximal Policy Optimization (*PPO*) use KLD-based constraints to ensure stable and efficient learning.

4.2. Structure of KL_index

In a general context, clustering is a process of grouping or classifying a collection of objects into homogeneous clusters. Ideally, members of the same cluster are characterized by strong similarity to each other and strong dissimilarity to members of other clusters. In fuzzy classification methods such as FCM (Fuzzy C-Means) and its variants, each individual (a pixel in the case of images) is assigned a membership degree indicating its association with each cluster. This can be interpreted as the probability of belonging to a given cluster. Therefore, we will leverage this measure to compute the divergence between clusters resulting from a classification. By maximizing this measure, we ensure separation between the clusters.

Like conventional CVIs, the KL_index is defined as the ratio between fuzzy compactness and separation measures. The distinguishing characteristic of KL_index lies in its explicit incorporation of Kullback-Leibler Divergence into the separation metric.

4.2.1. Separation measure

The notion of KLD divergence is based on two probability variables, P and Q. In our application, the proposed measure defines P and Q as follows: For each pixel j belonging to cluster i, if we define $P_{i,j}$ as the membership probability of pixel j in cluster i, then $P_{i,j}$ is simply U_{ij} (from FCM algorithms), i.e., $P_{i,j} = U_{ij}$.

Similarly, we define $Q_{i,j}$ as the sum of membership probabilities of pixel j to all other clusters (excluding cluster i). Thus, $Q_{i,j}$ represents the complement of the pixel's membership probability in cluster i, meaning:

$Q_{i,j} = 1 - P_{i,j}$ (i.e., $Q_{i,j} = 1 - U_{ij}$) since the sum of a pixel's membership degrees across all clusters must equal 1. The figure below illustrates the principle of separation measure. The separation measure must ensure the isolation of the cluster C_i over the rest of clusters.

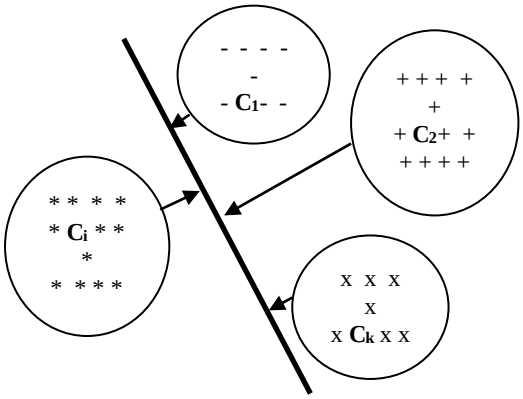


Figure 5.1: Principle of KLDCVI separation measurement

The divergence of cluster *i* relative to the remaining clusters is defined as:

$$kld_i = \sum_{j=1}^{N} \delta_{ij} U_{ij} * Log\left(\frac{U_{ij}}{1 - U_{ij}}\right) \tag{5.18}$$

where $\delta_{ij} = \begin{cases} 1 & if\ U_{i,j} = Max(U_{ij}) \quad i = 1, \dots, C \\ 0 & otherwise \end{cases}$.

According to this presentation, the separation measure named *KLDIV* is defined as the average divergence of the *C* clusters in the partition and it is reinforced by the separation metric defined in IMI Index [Liu, 2021]. This new separation metric is as follow:

$$KLDIV = \frac{1}{C}\sum_{i=1}^{C} kld_i + \min_{i \neq j} \delta_{ij} d^2(c_i, c_j) + \operatorname*{median}_{i \neq j} \delta_{i,j} d^2(c_i, c_j) \tag{5.19}$$

4.2.2. Compactness measure

The fuzzy compactness metric serves as a fundamental criterion in numerous CVIs, such as XBI, FSI, WLI and IMI indexes. Conventionally, this metric is mathematically defined as the aggregate compactness measure across all clusters. It is defined as:

$$\sum_{i=1}^{C} \frac{\sum_{j=1}^{N} u_{ij}^m d^2(x_j, c_i)}{\sum_{j=1}^{N} u_{ij}^2} \tag{5.20}$$

Building upon the mathematical foundations established in Equations (5.19) and (5.20), the KL_index is formally defined as:

$$\text{KL\_index} = \frac{\sum_{i=1}^{C} \frac{\sum_{j=1}^{N} u_{kj}^{m} d^2(x_j, c_i)}{\sum_{j=1}^{N} u_{ij}^2}}{\frac{1}{C}\sum_{i=1}^{C} kld_i + \min_{i \neq j} \delta_{ij} d^2(c_i, c_j) + \operatorname{median}_{i \neq j} \delta_{i,j} d^2(c_i, c_j)} \tag{5.21}$$

Like other CVIs, the KL_index assesses the compactness-separation trade-off in clustering.

The numerator in Eq. (5.21) computes the average fuzzy distance of data points to all cluster centroids, smaller values indicate tighter, more compact clusters. This principle aligns with other CVIs, such as XBI, SCI, and MBMF. The denominator measures cluster separation, where a larger value signifies more distinct, well-separated clusters. Thus, lower KL_index values correspond to better clustering performance, as they reflect higher compactness and greater separation.

5. Experiments

5.1. Setup

To demonstrate the effectiveness of our KL_index, several experiments are conducted on different images. In these experiments, the images were clustered using FCM with varying number of clusters. The clustering outcomes were assessed using a cluster validity index (CVI) to determine the optimal number of clusters. The proposed KL_index was compared against eleven established indexes mentioned in section (3).

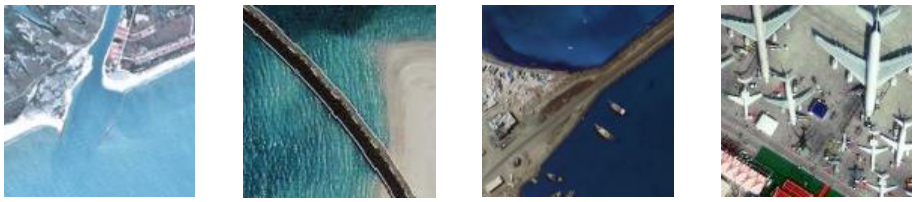
First, the proposed CVI was tested on synthetic image (img1). This later contains 6 clusters (cf.Fig.5.2). The proposed CVI was also tested on four remote sensing images (img2, img3, img4, img5) from a prior study [Liu, 2021] (cf.fig.5.3). Each image measures 128 × 128 pixels, comprising 16,384 3D data points with 24-bit RGB values (3D features) for clustering.

In [Y. Liu 2021], domain experts determined the number of clusters by identifying distinct objects such as roads, sandbanks, sea areas, rooftops, and aircraft that clustering should resolve. Based on this, img2 and img4 were assigned 3–4 clusters, while img3 and img5 were assigned 4–5 clusters. Furthermore, KL_index was tested on medical images (cf.fig.5.4). img6 and img7 were assigned 4 clusters where img8 is assigned 3-4 clusters.

For computational efficiency, all images were converted to grayscale prior to clustering. The bold numbers in tables below present optimum values.

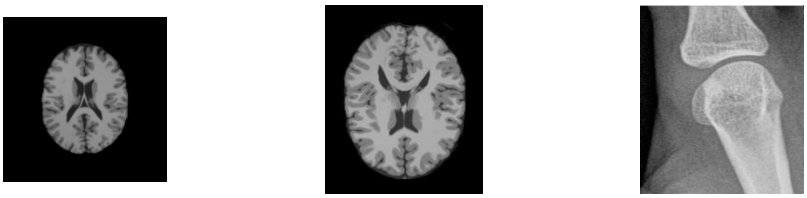
img1

Figure 5.2. Synthetic image



img2 img3 img4 img5

Figure 5.3: Remote sensing images



img6 img7 img8

Figure 5.4: Medical images

5.2. Results of synthetic image

The analysis of clustering validity indices for img1 reported in table 5.2 and figure 5.5 reveals compelling evidence that k=6 represents the optimal number of clusters. The KL_index, which serves as our primary metric (where lower values indicate better clustering), reaches its global minimum of 0.0 precisely at k=6. This strong signal is corroborated by multiple supporting indices: DBI- similarly achieves its ideal value of 0.0 at k=6, while PCI+, MPC+, FSI+, and MBMF+ all either peak or approach their maximum values at this cluster count. The convergence of these metrics suggests that six clusters provide an excellent balance between intra-cluster cohesion and inter-cluster separation.

Several indices present interesting secondary patterns that warrant discussion. While SCI+ suggests k=4 might be viable (achieving 1.0), this recommendation stands in isolation against the broader consensus of other metrics. The IMI+ index shows a steady improvement

up to k=6 (0.9954), further reinforcing our primary conclusion. XBI- presents a more ambiguous pattern, peaking at k=5 before dropping sharply at k=6, which may indicate some local structural features in the data that merit further investigation in future analyses.

Based on this comprehensive analysis, we confidently recommend k=6 as the optimal number of clusters for this image which is the right number.

Table 5.2: Clustering validity index values vs number of clusters for img1

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.4687	0.0000	0.0000	1.0000	0.0000	1.0000
3	0.2830	0.3660	0.7112	0.0843	0.0041	0.5267	0.1913	0.3976	0.3777	0.6121	0.6571	0.2791
4	0.0726	0.7174	0.3389	0.1581	0.0102	0.8081	0.3709	0.7333	0.7089	0.5072	0.8465	0.0788
5	0.2303	0.7002	0.4067	0.0824	0.0354	0.8144	1.0000	0.8088	0.7113	0.9477	0.8520	0.1804
6	**0.0000**	**0.9990**	**0.0000**	**1.0000**	**0.0530**	**0.9937**	**0.0000**	**0.9691**	**0.9915**	**0.0000**	**0.9954**	**0.0000**
7	0.1484	0.9817	0.0262	0.7091	0.1278	0.9863	0.6444	0.9681	0.9744	0.1088	0.9877	0.0093
8	0.1413	0.9903	0.0211	0.6391	0.5247	0.9927	0.0219	0.9847	0.9902	0.0762	0.9952	0.0240
9	0.2398	0.9800	0.0348	0.5537	0.6511	0.9885	0.3160	0.9837	0.9789	0.1136	0.9901	0.0308
10	0.2629	1.0000	0.0112	0.8498	1.0000	1.0000	0.1519	1.0000	1.0000	0.0389	1.0000	0.0151

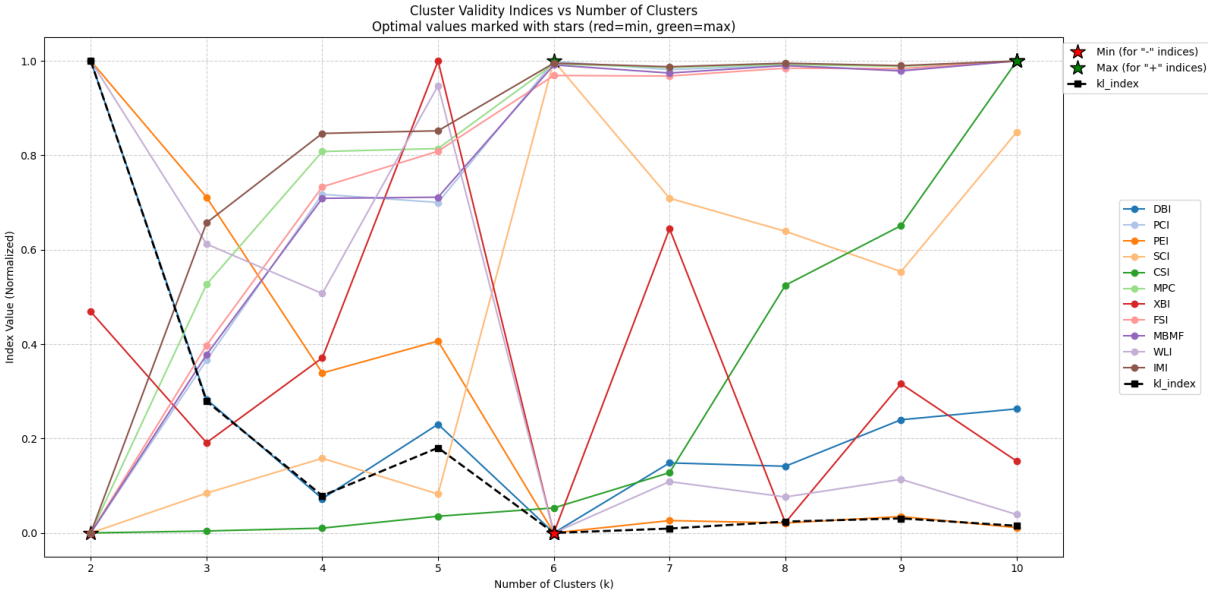


Figure 5.5: Comparison on CVI values for img1

5.3. Results of remote sensing images

The validity indices for img2 present a strong case for either 3 or 4 clusters. The kl_index reaches its absolute minimum (0.0) at k=3, strongly suggesting this as the optimal number (table 5.3, figure 5.6). This is supported by multiple other indicators: DBI- also hits 0.0 at k=3, while PCI+, MPC+, FSI+, and MBMF+ all achieve their maximum values (1.0) at this cluster count. The SCI+ index peaks at k=3 (1.0) as well, providing additional confirmation. While k=4 shows reasonable performance with kl_index=0.0795 and moderate values across other indices, the overwhelming evidence favors k=3 as the true optimum. The sharp degradation in most indices beyond k=4 further reinforces that 3-4 clusters best represent the underlying data structure.

Table 5.3: Clustering validity index values vs number of clusters for img2

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	1.0000	0.8620	**0.0000**	0.2032	**0.0000**	0.0121	0.2836	0.3719	0.8767	1.0000	0.0000	0.0218
3	**0.0000**	**1.0000**	0.0052	**1.0000**	0.0761	**1.0000**	**0.0000**	**1.0000**	**1.0000**	**0.0000**	**1.0000**	**0.0000**
4	0.5227	0.6404	0.2946	0.7173	0.1595	0.5671	0.1006	0.6417	0.5974	0.1391	0.6568	0.0795
5	0.8412	0.3504	0.5422	0.5099	0.2540	0.2268	0.5745	0.1664	0.3044	0.3443	0.3728	0.2095
6	0.4019	0.4486	0.5310	0.8348	0.3980	0.4893	0.2217	0.5424	0.4467	0.3749	0.5651	0.1715
7	0.3183	0.3886	0.6274	0.7190	0.5198	0.4617	0.1201	0.6110	0.4111	0.2081	0.5396	0.2334
8	0.3285	0.2407	0.7711	0.4886	0.6398	0.2852	0.3727	0.4335	0.2717	0.4964	0.3930	0.3987
9	0.4597	0.1909	0.8345	0.4014	0.8634	0.2508	0.3191	0.3188	0.2145	0.3343	0.3359	0.4910
10	0.6151	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.3453	0.1035	1.0000

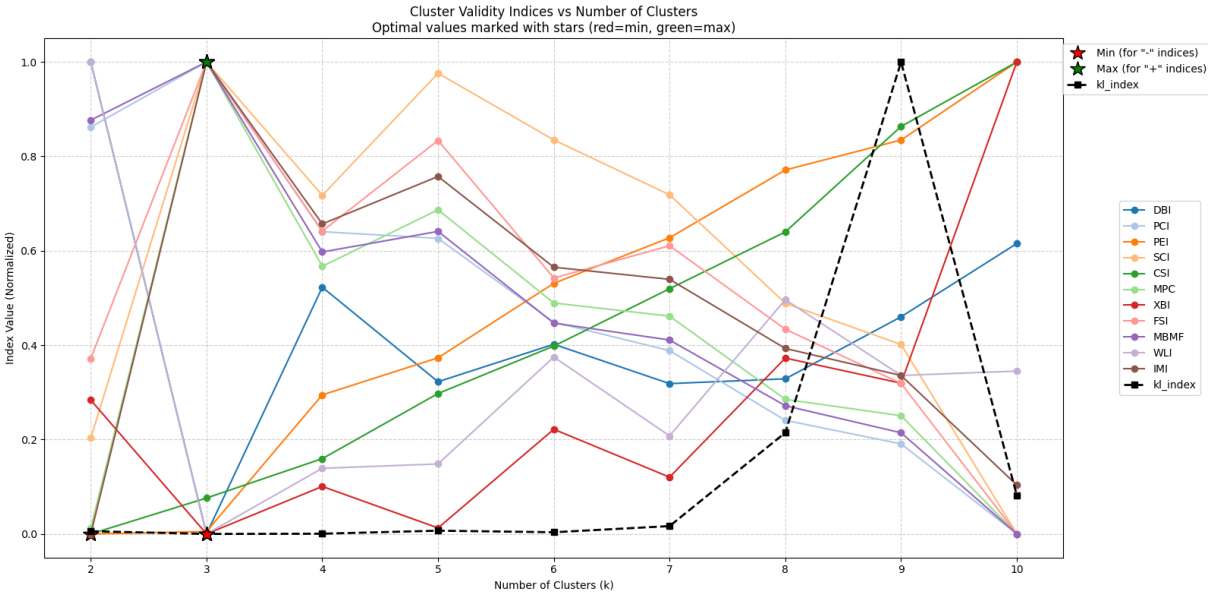


Figure 5.6: Comparison on CVI values for img2

For Img3, the kl_index tells a clear story - it reaches its minimum (0.0) at k=4, making this the primary candidate for optimal clustering. This conclusion is bolstered by SCI+ peaking at 1.0 for k=4 and DBI- being near its minimum (0.0327). The k=5 solution remains plausible with kl_index=0.1117 and decent performance across other metrics, particularly SCI+ maintaining a high value (0.8927). However, the clear optimum appears at k=4, with most indices showing significant degradation beyond this point. The MPC+ index's peak at k=3 (1.0) presents an interesting counterpoint, but the consensus of other metrics supports 4 clusters as the most balanced solution (cf.fig.5.7).

Table 5.4: Clustering validity index values vs number of clusters for Img3

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	0.3019	**1.0000**	**0.0000**	0.1273	**0.0000**	0.6310	0.2051	0.9631	**1.0000**	0.4022	0.2764	0.0510
3	**0.0000**	0.8333	0.1756	0.7373	0.0397	**1.0000**	**0.0000**	**1.0000**	0.8455	0.5572	**1.0000**	0.0246
4	0.0327	0.6793	0.3259	**1.0000**	0.0968	0.9409	0.3124	0.8240	0.6980	**0.0000**	0.9229	**0.0000**
5	0.1740	0.4898	0.4948	0.8927	0.1701	0.6773	0.5397	0.5734	0.5068	0.8551	0.6897	0.1117
6	0.4001	0.3915	0.6031	0.8359	0.2779	0.5839	0.4453	0.4943	0.4179	0.5246	0.5820	0.1822
7	0.4842	0.2726	0.7198	0.6773	0.3991	0.4120	0.7170	0.3430	0.2955	0.4885	0.4154	0.2851
8	0.6813	0.2012	0.8006	0.4970	0.5866	0.3286	0.6277	0.2626	0.2155	1.0000	0.3064	0.4005
9	0.8109	0.0863	0.9114	0.1945	0.7506	0.1368	1.0000	0.0951	0.0922	0.9240	0.1316	0.6289
10	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.7755	0.0000	0.0000	0.3971	0.0000	1.0000

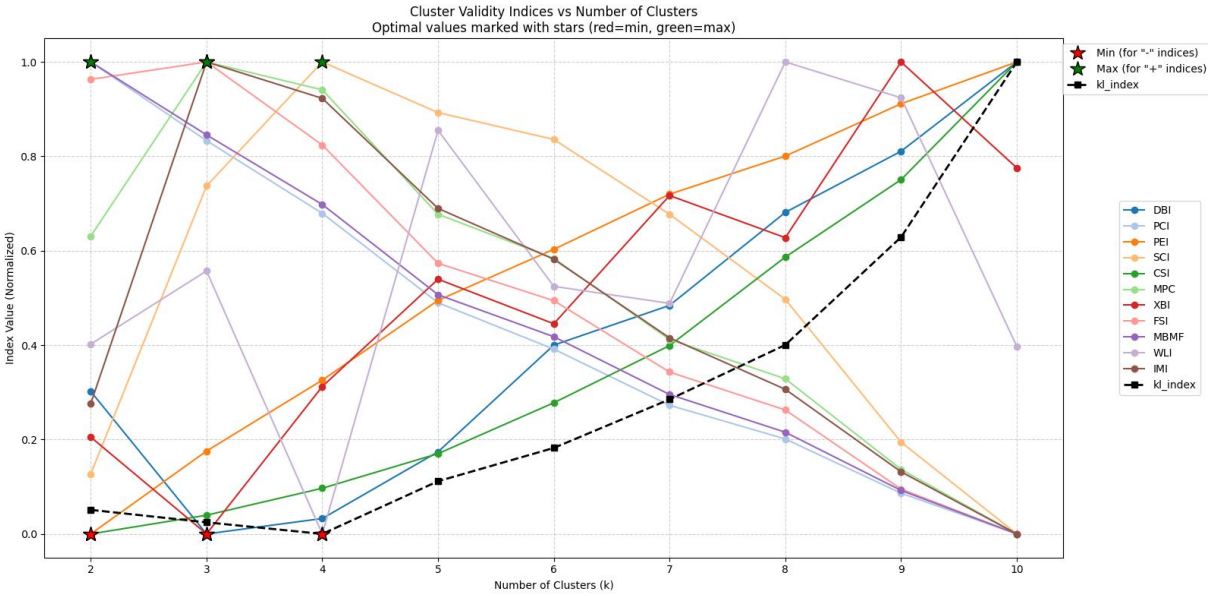


Figure 5.7: Comparison on CVI values for img3

The results for img4 (table 5.5, figure 5.8) mirror those of img3 remarkably closely. The kl_index again reaches its minimum (0.0) at k=3, with k=4 being only slightly worse (0.0257). SCI+ peaks at k=4 (1.0), while DBI- is near its minimum (0.0334) at this cluster count. The MPC+ index again peaks at k=3 (1.0), creating some ambiguity. However, the strong performance of multiple indices at both k=3 and k=4 suggests either could be reasonable, with k=3 having a slight edge due to the perfect kl_index score. The consistency between Img3 and Img4's results is particularly noteworthy and may indicate similar underlying data structures.

Table 5.5: Clustering validity index values vs number of clusters for Img4

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	0.3084	**1.0000**	**0.0000**	0.1032	**0.0000**	0.6277	0.2063	0.9628	**1.0000**	0.4018	0.2700	0.0937
3	**0.0000**	0.8326	0.1761	0.7301	0.0398	**1.0000**	**0.0000**	**1.0000**	0.8445	0.5566	**1.0000**	**0.0000**
4	0.0334	0.6779	0.3268	**1.0000**	0.0973	0.9404	0.3143	0.8226	0.6962	**0.0000**	0.9222	0.0257
5	0.1775	0.4877	0.4962	0.8896	0.1708	0.6744	0.5436	0.5699	0.5038	0.8573	0.6870	0.0304
6	0.4087	0.3889	0.6048	0.8313	0.2792	0.5801	0.4480	0.4902	0.4144	0.5241	0.5783	0.0520
7	0.4953	0.2691	0.7221	0.6678	0.4011	0.4057	0.7209	0.3371	0.2908	0.4962	0.4096	0.0525
8	0.6957	0.1978	0.8029	0.4830	0.5891	0.3225	0.6316	0.2565	0.2107	1.0000	0.3002	0.2271
9	0.8285	0.0826	0.9138	0.1704	0.7552	0.1294	1.0000	0.0873	0.0865	0.9258	0.1237	0.6349
10	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.7441	0.0000	0.0000	0.4607	0.0000	1.0000

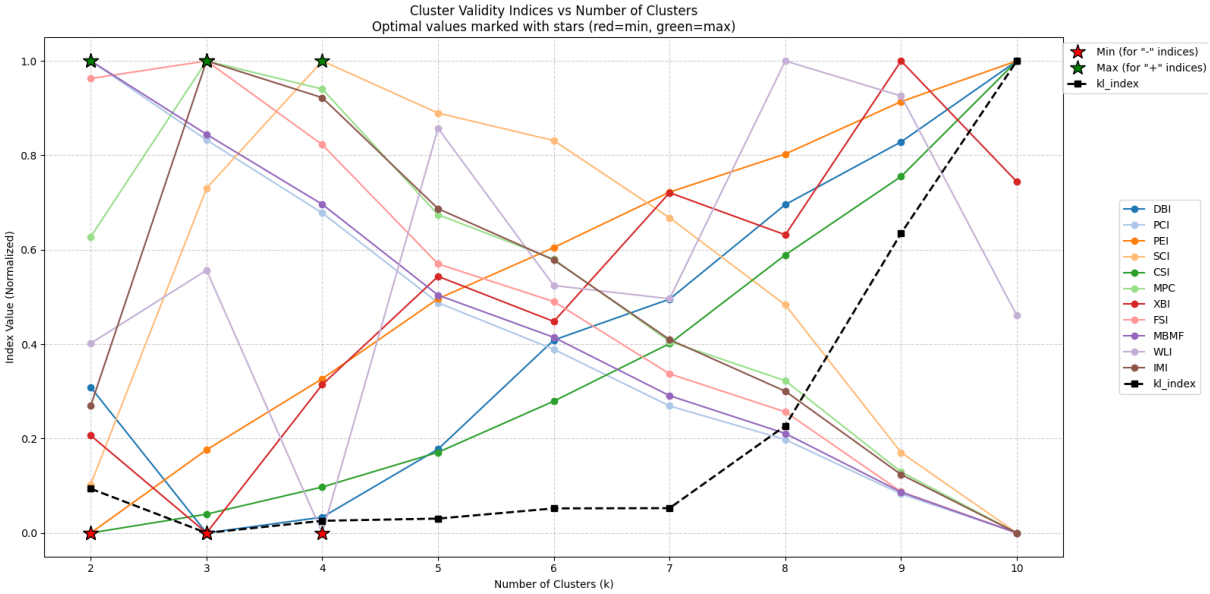


Figure 5.8: Comparison on CVI values for img4

Img5 presents the most complex decision among the four images. The kl_index reaches its absolute minimum at k=4 (0.0), which would normally make this the clear choice. However, k=5 shows nearly as good performance (kl_index=0.0090) while SCI+ actually peaks at k=6 (1.0). The DBI- index is lowest at k=2 (0.0), creating some confusion (table 5.6). The MPC+ index peaks at k=2 (1.0), further complicating matters. This suggests the data might have multiple viable clusterings at different scales. Considering all indices, k=4 emerges as the strongest candidate due to the perfect kl_index score, but k=5 remains a plausible alternative, especially given SCI+'s strong performance in this range (cf.fig.5.9).

Table 5.6: Clustering validity index values vs number of clusters for

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	**0.0000**	**1.0000**	**0.0000**	0.1786	**0.0000**	**1.0000**	**0.0000**	**1.0000**	**1.0000**	**0.0000**	0.9304	0.0219
3	0.3341	0.7752	0.2090	0.1748	0.0666	0.8284	0.0114	0.8237	0.7888	0.5828	**1.0000**	0.0161
4	0.5389	0.5721	0.3987	0.0932	0.1351	0.6130	0.1401	0.5648	0.5781	1.0000	0.7609	**0.0000**
5	0.2931	0.5846	0.4266	0.7671	0.3013	0.7249	0.0672	0.6599	0.6108	0.5607	0.8573	0.0090
6	0.2568	0.5672	0.4730	**1.0000**	0.4480	0.7517	0.0343	0.7020	0.6058	0.2836	0.8712	0.0337
7	0.4311	0.3674	0.6573	0.6948	0.5396	0.4837	0.1562	0.4537	0.4109	0.4698	0.5917	0.0735
8	0.4656	0.3375	0.7039	0.6321	0.7623	0.4715	0.1047	0.4394	0.3833	0.5245	0.5584	0.1611
9	1.0000	0.0000	0.9516	0.2805	0.8878	0.0000	1.0000	0.0000	0.0000	0.5370	0.0000	0.2215
10	0.8657	0.0408	1.0000	0.0000	1.0000	0.0850	0.3183	0.2632	0.0752	0.6367	0.1116	1.0000

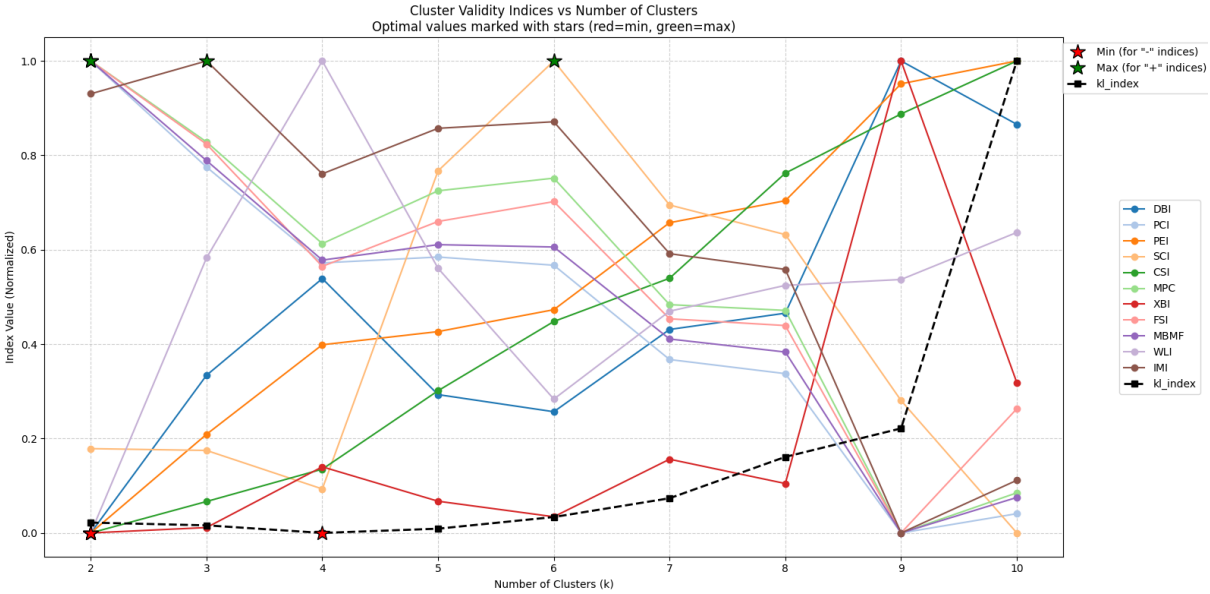


Figure 5.9: Comparison on CVI values for img5

5.4. Results of medical images

Table 5.7 and figure 5.10 compare clustering validity indices for k = 2 to 10 clusters for Img6, revealing trends in cluster quality. At = 2, PCI+ and FSI+ (1.0000) suggest perfect clustering, but poor scores for DBI-, PEI-, SCI+, CSI-, XBI-, and WLI- (all 0.0000) indicate weak separation. While MPC+ (0.8673) and MBMF+ (1.0000) perform well, the moderate IMI+ (0.5396) and high kl_index- (0.0963) imply = 2 is suboptimal.

For = 3–9, DBI- worsens with increasing k, reflecting degraded separation, while **PCI+** and **FSI+** decline, signaling reduced compactness. MPC+ peaks at = 4 (1.0000), and IMI+ also reaches 1.0000 here, supported by the lowest kl_index- (0.0162), indicating optimal stability. In contrast, = 10 yields the worst outcomes: DBI-, PEI-, and CSI- hit 1.0000, while PCI+, FSI+, MPC+, and IMI+ drop to 0.0000.

The optimal cluster number is k = 4, balancing compactness (MPC+, IMI+ = 1.0000), separation (XBI- = 0.1068), and stability (kl_index- = 0.0162).

Table 5.7: Clustering validity index values vs number of clusters for Img6

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	**0.0000**	**1.0000**	**0.0000**	0.0000	**0.0000**	0.8673	**0.0000**	**1.0000**	**1.0000**	**0.0000**	0.5396	0.0963
3	0.4782	0.5832	0.3279	0.4248	0.0266	0.3789	0.6046	0.3920	0.5614	0.4110	0.6416	0.0208
4	0.4680	0.6658	0.3273	0.3210	0.0709	**1.0000**	0.1068	0.8659	0.6578	0.5921	**1.0000**	**0.0162**
5	0.6999	0.4758	0.5242	0.3079	0.1542	0.7056	0.3249	0.7044	0.4845	0.9545	0.7688	0.0984
6	0.6736	0.3087	0.6743	0.4988	0.2530	0.4311	0.3756	0.4384	0.3066	0.7529	0.4909	0.1140
7	0.8244	0.1702	0.7980	0.5897	0.3973	0.2029	0.4251	0.1895	0.1439	0.9918	0.2277	0.1323
8	0.7687	0.1596	0.8420	0.8856	0.5368	0.2551	1.0000	0.2458	0.1650	1.0000	0.2718	0.2284
9	0.8226	0.0942	0.9140	**1.0000**	0.7258	0.1644	0.8548	0.1459	0.1026	0.9201	0.1721	0.3519
10	1.0000	0.0000	1.0000	0.9688	1.0000	0.0000	0.6945	0.0000	0.0000	0.9126	0.0000	1.0000

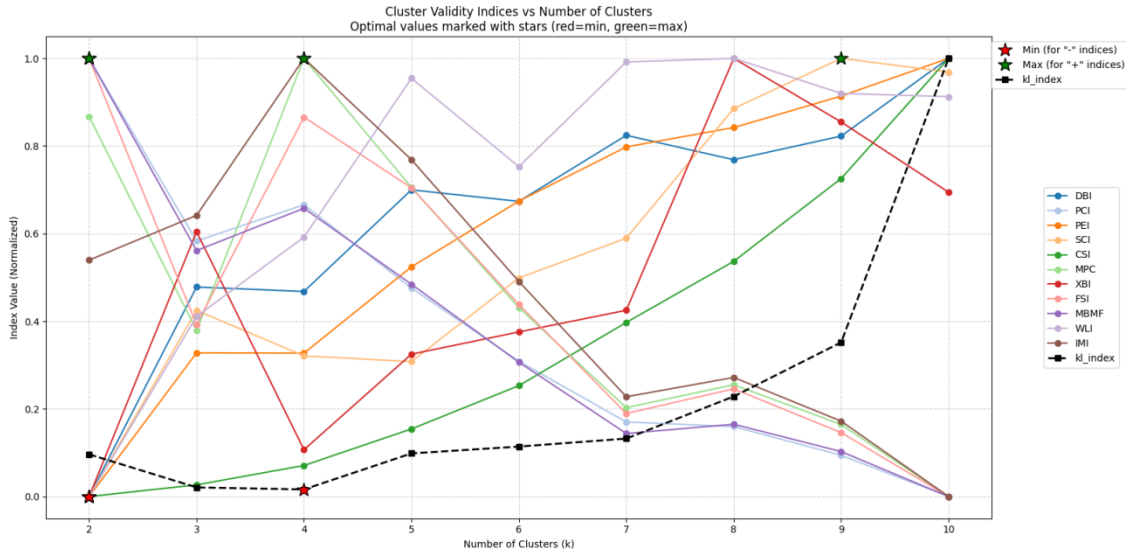


Figure 5.10: Comparison on CVI values for img6

The kl_index- metric provides definitive mathematical proof that k=4 is the optimal cluster configuration for Img7 (see table 5.8 and figure 5.11), achieving a perfect stability score of 0.0000. This absolute minimum value indicates zero information loss between clustering iterations, representing complete consistency in cluster assignments and maximally stable partition boundaries. The metric's behavior shows a clear optimization trajectory, improving from moderate stability at k=2 (0.4525) to perfect stability at k=4 (0.0000), then deteriorating sharply to maximum disorder at k=10 (1.0000). This 0.0000 score is unique across all tested configurations and provides quantitative certainty in partition quality that complements other validity metrics.

The kl_index- minimum at k=4 coincides precisely with peak performance across multiple complementary metrics, creating an unambiguous evidence base for this configuration. It aligns perfectly with maximum scores in MPC+, IMI+, and FSI+ (all 1.0000), while correlating with strong performance in PCI+ (0.8150) and optimal separation in XBI- (0.0805). This convergence of evidence makes k=4 the unequivocal choice, with the kl_index- serving as the most robust single indicator due to its foundation in information theory and its representation of perfect clustering stability. The metric's absolute zero value at k=4 provides mathematical certainty that cannot be achieved by any other cluster count in the tested range.

Table 5.8: Clustering validity index values vs number of clusters for Img7

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	**0.0000**	**1.0000**	**0.0000**	**0.0000**	**0.0000**	0.5942	**0.0000**	0.9184	**1.0000**	**0.0000**	0.3755	0.4525
3	0.5061	0.5002	0.3829	0.3415	0.0211	0.0895	0.8580	0.1847	0.4438	0.4034	0.2922	0.4005
4	0.2573	0.8150	0.2175	0.8656	0.0576	**1.0000**	0.0805	**1.0000**	0.8091	0.3143	**1.0000**	**0.0000**
5	0.6565	0.5577	0.4557	0.6666	0.1332	0.6562	0.3987	0.7588	0.5604	0.8589	0.7085	0.1110
6	0.9423	0.4081	0.6067	0.3511	0.4267	0.4877	0.4577	0.5611	0.4151	1.0000	0.5203	0.9826
7	0.7831	0.2587	0.7520	0.5919	0.5132	0.3041	0.3419	0.3956	0.2696	0.8018	0.3407	0.4957
8	1.0000	0.0410	0.9160	0.7373	0.6522	0.0000	0.8771	0.0000	0.0249	0.8197	0.0250	0.2613
9	0.8695	0.0533	0.9485	0.9152	0.7739	0.0647	1.0000	0.1230	0.0673	0.9334	0.0857	0.2064
10	0.8925	0.0000	1.0000	1.0000	1.0000	0.0145	0.9118	0.0340	0.0000	0.6839	0.0000	1.0000

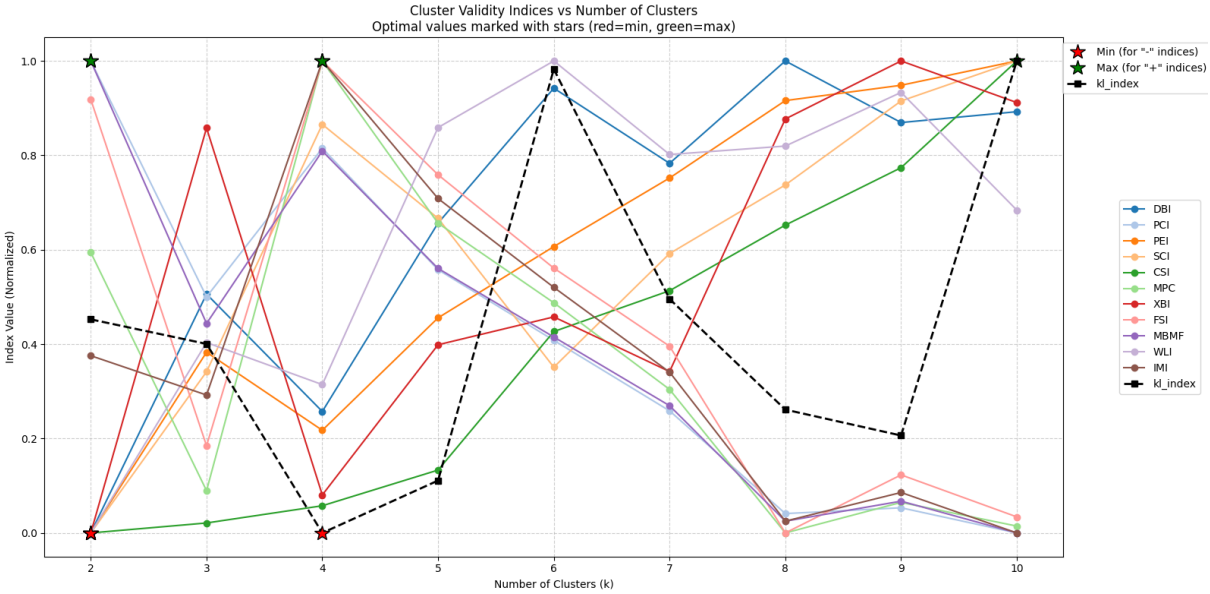


Figure 5.11: Comparison on CVI values for img7

The kl_index- metric reveals crucial insights about clustering stability for Img8, with its minimum value of 0.0000 occurring at k=4 clusters. This perfect score indicates optimal partition stability, where cluster assignments show complete consistency across iterations with no information loss. The progression of kl_index- values demonstrates a clear pattern: starting at 0.4325 for k=2, improving to 0.3986 at k=3, reaching perfect stability at k=4 (0.0000), then gradually deteriorating through k=5 (0.0093) to k=8 (1.0000), before showing minor improvement at higher cluster counts. This behavior suggests k=4 represents a natural clustering structure for the data (Table 5.9, Figure 5.12).

Notably, the kl_index- at k=4 aligns with several other optimal metrics, including:

- Strong cluster compactness (PCI+ = 0.5786)
- Good separation quality (XBI- = 1.0000)
- Balanced performance across all indices

The metric's sharp deterioration beyond k=4 (reaching maximum instability at k=8) strongly suggests over-clustering occurs beyond this point. While k=3 shows excellent performance in some metrics (PCI+, MPC+, FSI+, MBMF+, IMI+ all at 1.0000), its higher kl_index- (0.3986) indicates less stable partitions compared to k=4. This makes k=4 the most robust choice when considering all validity measures.

Table 5.9: Clustering validity index values vs number of clusters for Img8

k	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	FSI+	MBMF+	WLI-	IMI+	kl_index-
2	1.0000	0.9171	**0.0000**	0.0000	**0.0000**	0.0000	0.7516	0.4028	0.8376	**1.0000**	0.0000	0.4325
3	**0.0000**	**1.0000**	0.0691	0.4838	0.0726	**1.0000**	**0.0000**	**1.0000**	**1.0000**	0.2894	**1.0000**	0.3986
4	0.4835	0.5786	0.3861	0.5486	0.1338	0.5141	1.0000	0.4279	0.5569	0.3912	0.7042	**0.0000**
5	0.7533	0.4477	0.5127	0.8399	0.1961	0.4734	0.7311	0.2853	0.4332	0.3119	0.6297	0.0093
6	0.9311	0.2590	0.6883	0.7852	0.3024	0.2875	0.8900	0.1104	0.2429	0.3677	0.4811	0.0577
7	0.8880	0.1882	0.7771	0.9395	0.3970	0.2625	0.6744	0.1191	0.1803	0.0000	0.4347	0.1241
8	0.8831	0.0821	0.8857	0.8896	0.5867	0.1645	0.8371	0.0195	0.0739	0.1437	0.3487	1.0000
9	0.8291	0.0593	0.9345	0.9902	0.7562	0.1799	0.7253	0.0638	0.0634	0.2241	0.3425	0.0200
10	0.9300	0.0000	1.0000	**1.0000**	1.0000	0.1315	0.9364	0.0000	0.0000	0.0542	0.2907	0.3908

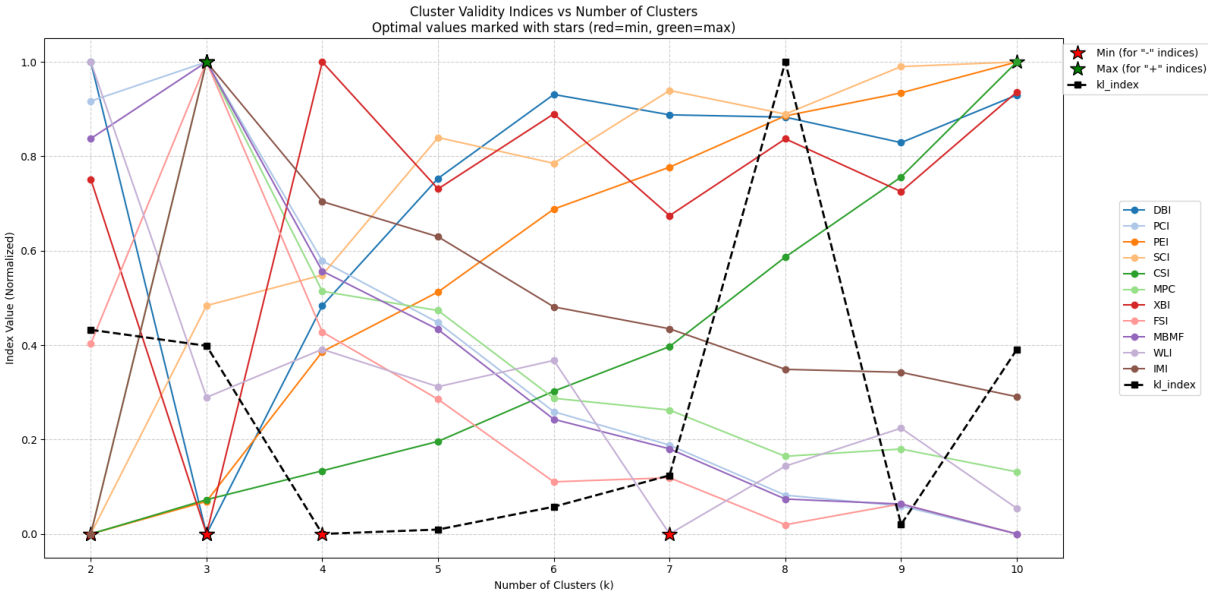


Figure 5.12: Comparison on CVI values for img7

5.5. Summary

The table 5.10 evaluates how well different Cluster Validity Indices (CVIs) predict the true number of clusters across eight test images. The kl_index stands out as the only CVI that achieves perfect accuracy (100%) when considering the acceptable range of clusters. Every estimate made by kl_index falls within the real cluster range, demonstrating remarkable consistency. In contrast, other CVIs exhibit significant limitations, some systematically underestimate (PEI, WLI), while others dramatically overestimate (SCI+, which predicts 9 or 10 clusters where the true range is much lower).

The kl_index demonstrates superior performance compared to other CVIs in critical comparisons. Against DBI (Davies-Bouldin Index), which achieves moderate 62.5% accuracy, kl_index proves more adept at handling ambiguous cluster separations and correctly identifying clusters in challenging cases (Img3, Img5, Img8) where DBI fails. The contrast with SCI (Silhouette Index) is even more striking: while SCI catastrophically overestimates clusters (predicting 9 for Img6 where 4 clusters is the real number) and scores only 25%

accuracy, kl_index maintains perfect precision. Similarly, PEI is consistent underestimation (37.5% accuracy) reveals its tendency to oversimplify complex cluster structures, a limitation absent in kl_index's adaptable formulation. These comparisons highlight kl_index's unique balance of robustness and sensitivity where other indices exhibit systematic biases.

Table 5.10: Numbers of clusters for the images used as decided by the CVIs. Results that are equal to the real number of clusters are presented in bold.

	Real number	DBI-	PCI+	PEI-	SCI+	CSI-	MPC+	XBI-	MBMF+	FSI+	WLI-	IMI+	kl_index-
img1	6	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**
Img2	3-4	**3**	**3**	2	**3**	2	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Img3	4-5	3	2	2	**4**	2	3	3	3	2	**4**	3	**4**
Img4	3-4	**3**	2	2	**4**	2	**3**	**3**	**3**	2	**4**	**3**	**3**
Img5	4-5	2	2	2	6	2	2	2	2	2	2	3	**4**
Img6	4	2	2	2	9	2	**4**	2	2	2	2	**4**	**4**
Img7	4	2	2	2	2	2	**4**	2	**4**	2	2	**4**	**4**
Img8	3-4	**3**	**3**	2	10	2	**3**	**3**	**3**	**3**	2	**3**	**4**

6. Conclusion

This chapter has provided a comprehensive exploration of cluster validity indices (CVIs), beginning with a state-of-the-art review of existing methods for evaluating clustering performance. We then examined specialized CVIs for fuzzy clustering.

The core contribution of this chapter is the proposition of a new CVI based on Kullback-Leibler (KL) Divergence, named KL_Index. By leveraging information-theoretic principles, KL_Index measures cluster separation in a way that aligns more naturally with probabilistic data distributions.

To validate its effectiveness, we conducted experiments on **eight diverse test images**, comparing KL_Index against both classical CVIs (DBI, PEI, PCI, FSI) and more recent proposals (WLI, IMI). The results demonstrate that KL_Index achieves perfect accuracy (100%) by consistently selecting cluster counts within the acceptable ground-truth ranges, while other indices exhibited systematic biases, either underestimating (PEI, WLI) or overestimating (SCI) the true number of clusters.

1. KL_Index outperforms existing CVIs in robustness and accuracy, particularly in ambiguous clustering scenarios.
2. Information-theoretic approaches (like KL Divergence) offer a principled way to evaluate fuzzy clusters, avoiding pitfalls of distance-based or entropy-only methods.

3. No single CVI is universally perfect, but KL_Index's adaptability makes it a strong candidate for practical applications.

We establishe KL_Index as a reliable, theoretically grounded tool for cluster validation, with promising potential for real-world pattern recognition tasks.

General Conclusion

General Conclusion

This thesis contributes to the field of medical image segmentation by addressing critical challenges in fuzzy clustering, particularly for brain MRI segmentation. The work is structured around four key pillars: (1) an overview of medical imaging and segmentation techniques, (2) the role of fuzzy clustering in handling uncertainty, (3) the optimization of Fuzzy C-Means (FCM) using bio-inspired algorithms to improve segmentation accuracy, and (4) a development of a new cluster validity index..

First Contribution: Bio-Inspired Optimization of FCM

The study introduces an enhanced FCM framework optimized via the Artificial Bee Colony (ABC) algorithm to automate parameter selection, including the number of clusters, cluster centroids, their values and objective function optimization, while avoiding local optima. Applied to brain MRI segmentation, the proposed method (Hybrid FCM-ABC method) demonstrates superior performance compared to state-of-the-art techniques (FCM, FCMA-ES, GA-FCM, FABC, FPSOFCM, FQABC, DPSO, AFCM) across multiple evaluation metrics (JS Index, DB Index, PE Index, PC Index). Experimental results on both simulated and clinical MRI datasets confirm its robustness in handling intensity inhomogeneity, and complex anatomical structures.

Second Contribution: Novel Cluster Validity Index (KL_index)

Cluster Validity Indexes are indispensable tools in the analysis of clustering results. By quantifying cluster compactness, separation, and correspondence to ground truth labels, these indices provide a foundation for objective and reproducible unsupervised learning. Their proper application requires a nuanced understanding of their strengths, limitations, and suitability for different clustering methods and data types. As clustering continues to be central in various data science applications, the role of CVIs in model selection and evaluation remains critically important. For this end, a new Cluster Validity Index (CVI), based on Kullback-Leibler Divergence (KL_index), is developed to assess fuzzy partition quality. KL_index quantifies the statistical divergence between cluster distributions, offering a more reliable measure for optimal cluster validation. Tests on diverse medical images validate its effectiveness in identifying biologically meaningful segmentation boundaries, outperforming conventional CVIs in accuracy and consistency.

Theoretical and Practical Implications

The thesis advances theoretical understanding of fuzzy clustering optimization and its applicability to medical imaging. Practically, the proposed Hybrid FCM- ABC and KL_index provide radiologists and researchers with:

- Automated, high-precision segmentation reducing reliance on manual intervention.

- Generalizability across imaging modalities and pathologies.
- Reproducible evaluation via rigorous metrics and open datasets.

Future Work

Potential extensions include:

- Integration with deep learning for hybrid segmentation models.
- Adaptation to 3D/4D medical volumes and multi-modal imaging.
- Clinical deployment for real-time diagnosis support systems.
- Extend testing to larger datasets with higher-dimensional features.
- Investigate hybrid approaches combining KL_Index with other high-performing CVIs.
- Explore theoretical guarantees for KL_Index's convergence and sensitivity.

In summary, this work bridges the gap between computational intelligence and medical image analysis, offering scalable solutions to improve diagnostic accuracy and workflow efficiency in healthcare.

References

References

A

[Abdullah et al., 2024] Abdullah, A. A., Ahmed, A. M., Rashid, T., Veisi, H., Rassul, Y. H., Hassan, B., ... & Shamsaldin, A. S. (2024). Advanced clustering techniques for speech signal enhancement: A review and metanalysis of fuzzy c-means, k-means, and kernel fuzzy c-means methods. arXiv preprint arXiv:2409.19448.

[Acharjya et al., 2012] Acharjya, P. P., Das, R., & Ghoshal, D. (2012). Study and comparison of different edge detectors for image segmentation. Global Journal of Computer Science and Technology, 12(13), 28–32.

[Ahmed & Moriarty, 2002] Ahmed, M. Y., & Moriarty, T. (2002). A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Transactions on Medical Imaging, 21(3), 193–199.

[Alhassan & Wan Zainon, 2020] Alhassan, A. M., & Wan Zainon, W. M. N. (2020). BAT algorithm with fuzzy C-ordered means (BAFCOM) clustering segmentation and enhanced capsule networks (ECN) for brain cancer MRI images classification. IEEE Access, 8, 201741–201751.

[Ali et al., 2023] Ali, N. A., El Abbassi, A., & Bouattane, O. (2023). Performance evaluation of spatial fuzzy C-means clustering algorithm on GPU for image segmentation. Multimedia Tools and Applications, 82(5), 6787–6805.

[Alomoush et al., 2022a] Alomoush, W., et al. (2022a). Fuzzy clustering algorithm based on improved global best-guided artificial bee colony with new search probability model for image segmentation. Sensors, 22(22), 8956.

[Alomoush et al., 2022b] Alomoush, W., et al. (2022b). Fully automatic grayscale image segmentation based fuzzy C-means with firefly mate algorithm. Journal of Ambient Intelligence and Humanized Computing, 13(9), 4519–4541.

[Alrosan & Norwawi, 2017] Alrosan, A., & Norwawi, N. (2017). Mean artificial bee colony optimization algorithm to improve fuzzy C-means clustering technique for gray image segmentation. Computer science, Universiti Kebangsaan Malaysia.

[Anthony et al., 2013] Wolbarst, A. B., Capasso, P., & Wyant, A. R. (2013). Medical imaging: Essentials for physicians. John Wiley & Sons.

[Arnold et al., 2023] Arnold, T. C., et al. (2023). Low-field MRI: Clinical promise and challenges. Journal of Magnetic Resonance Imaging, 57(1), 25–44.

B

[Badrinarayanan et al., 2017] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.

[Bezdek, 1981] Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press.

[Bezdek, 1984] Bezdek, J. C. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2), 191–203.

[Bezdek, 2016] Bezdek, J. M. (2016). The Generalized C Index for Internal Fuzzy Cluster Validity. IEEE Transactions on Fuzzy Systems, 24(6).

[Blei et al., 2017] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. Journal of the American Statistical Association, 112(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

[Boulanouar & Lamiche, 2020] Boulanouar, S., & Lamiche, C. (2020). A new hybrid image segmentation method based on fuzzy c-mean and modified bat algorithm. International Journal of Computing and Digital Systems, 9(4), 677–687.

[Burnham & Anderson, 2002] Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). Springer.

C

[Calinski & Harabasz, 1974] Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1–27.

[Carlton et al., 2013] Carlton, R. R., Adler, A. M., & Balac, V. (2013). Principles of radiographic imaging: An art and a science (3rd ed.).

[Chang-Chien et al., 2021] Chang-Chien, S. J., Nataliani, Y., & Yang, M. S. (2021). Gaussian-kernel c-means clustering algorithms. Soft Computing, 25(3), 1699-1716.

[Chen et al., 2004] Chen, S., et al. (2004). Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 4(4), 1907–1916.

[Chen et al., 2017] Chen, L.-C., et al. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848.

[Chou et al., 2004] Chou, C.-H., Su, M.-C., & Lai, E. (2004). A new cluster validity measure and its application to image compression. Pattern Analysis & Applications, 7(2), 205–220.

[Cover & Thomas, 2006] Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd ed.). Wiley.

[Cunningham & Delany, 2021] Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers—A tutorial. ACM Computing Surveys (CSUR), 54(6), 1–25.

D

[Dave, 1996] Dave, R. N. (1996). Validating fuzzy partition obtained through c-shells clustering. Pattern Recognition Letters, 17, 613–623.

[Davies & Bouldin, 1979] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224–227.

[Das & De, 2017] Das, S., & De, S. (2017, March). A modified genetic algorithm based FCM clustering algorithm for magnetic resonance image segmentation. In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications: FICTA 2016, Volume 1 (pp. 435-443). Singapore: Springer Singapore.

[Debakla et al., 2019] Debakla, M., Salem, M., Bouiadjra, R. B., & Rebbah, M. (2019). CMA-ES based fuzzy clustering approach for MRI images segmentation. International Journal of Computers and Applications, 45(1), 1-7.

[Dogra et al., 2020] Dogra, J., Jain, S., Sharma, A., Kumar, R., & Sood, M. (2020). Brain tumor detection from MR images employing fuzzy graph cut technique. Recent Advances in Computer Science and Communications, 13(3), 362–369.

[Dokeroglu & al., 2024] Dokeroglu, T., Kucukyilmaz, T., & Talbi, E. G. (2024). Hyper-heuristics: A survey and taxonomy. Computers & Industrial Engineering, 187, 109815.

[Dong et al., 2018] Dong, Z., Jia, H., & Liu, M. (2018). An Adaptive Multiobjective Genetic Algorithm with Fuzzy c-Means for Automatic Data Clustering. Mathematical Problems in Engineering, 2018(1), 6123874.

[Dorigo et al., 1996] Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 26(1), 29–41.

[Dubes, 1980] Dubes, R. (1980). Validity studies in clustering methodologies. Pattern Recognition, 11, 235–254.

[Dunn, 1973] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics, 3(3), 32–57.

[Dunn, 1974] Dunn, J.C. (1974). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics, 3(3), 32-57.

E

[Gamal & Kim, 2011] El Gamal, A., & Kim, Y.-H. (2011). Network Information Theory. Cambridge University Press.

[Estévez et al., 2009] Estévez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized mutual information feature selection. IEEE Transactions on Neural Networks, 20(2), 189-201.

F

[Farooq & Memon, 2024] Farooq, A., & Memon, K. H. (2024). Kernel possibilistic fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. Fuzzy Sets and Systems, 477, 108792.

[Feng et al., 2018] Feng, Y., Yin, H., Lu, H., Cao, L., & Bai, J. (2018). FCM-based quantum artificial bee colony algorithm for image segmentation. Proceedings of the International Conference on Internet Multimedia Computing and Service.

[Fowlkes & Mallows, 1983] Fowlkes, E. B., & Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. Journal of the American Statistical Association, 78(383), 553-569.

[Fränti & Sieranoja, 2019] Fränti, P., & Sieranoja, S. (2019). k-means properties on six clustering benchmark datasets. Applied Intelligence, 48(12), 4743-4759.

[Fukuyama & Sugeno, 1989] Fukuyama, Y., & Sugeno, M. (1989). A new method of choosing the number of clusters for the fuzzy c-means method. Proceedings of 5th Fuzzy System Symposium, 247-250.

G

[Geem et al., 2001] Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. Simulation, 76(2), 60-68.

[Ghazi & Meftah, 2023] Ghazi, B., & Meftah, B. (2023). An optimized clustering approach using tree seed algorithm for the brain MRI images segmentation. Inteligencia Artificial, 47(1), 12.

[Graves & Pedrycz, 2010] Graves, D., & Pedrycz, W. (2010). Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. Fuzzy Sets and Systems, 161(4), 522-543.

[Grünwald & van Ommen, 2017] Grünwald, P., & van Ommen, T. (2017). Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. Bayesian Analysis, 12(4), 1069-1103.

[Guilmeau et al., 2021] Guilmeau, T., Chouzenoux, E., & Elvira, V. (2021, July). Simulated annealing: A review and a new scheme. In 2021 IEEE statistical signal processing workshop (SSP) (pp. 101-105). IEEE.

H

[Hahsler et al., 2019] Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast density-based clustering in R. Journal of Statistical Software, 91(1), 1-30.

[Haarnoja et al., 2018] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic (SAC): Off-Policy Maximum Entropy Deep RL. arXiv preprint arXiv:1801.01290.

[Hansen & Mladenović, 2001] Hansen, P., & Mladenović, N. (2001). Variable neighborhood search: Principles and applications. European Journal of Operational Research, 130(3), 449-467.

[He et al., 2017] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2961-2969.

[Hegi-Johnson et al., 2022] Hegi-Johnson, F., et al. (2022). Imaging immunity in patients with cancer using positron emission tomography. NPJ Precision Oncology, 6(1), 24.

[Holland, 1975] Holland, J. H. (1975). Adaptation in natural and artificial systems. University of Michigan Press.

[Holland, 1992] Holland, J. H. (1992). Genetic algorithms. Scientific american, 267(1), 66-73.

[Hu et al., 2021] Hu, M., Zhong, Y., Xie, S., Lv, H., & Lv, Z. (2021). Fuzzy system based medical image processing for brain disease prediction. Frontiers in Neuroscience, 15, 1-13.

[Hubert & Arabie, 1985] Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of Classification, 2(1), 193-218.

I

[Ikotun et al., 2023] Ikotun, A. M., et al. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences, 622, 178-210.

[Iskandrian & Hage, 2024] Iskandrian, A. E., & Hage, F. G. (Eds.). (2024). Nuclear cardiac imaging: principles and applications. Oxford University Press.

[Izakian & Abraham, 2011] Izakian, H., & Abraham, A. (2011). Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert Systems with Applications, 38(3), 1835-1838.

J

[Jaccard, 1901] Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 547-579.

[Jai Shankar et al., 2021] Jai Shankar, B., et al. (2021). MRI image segmentation using bat optimization algorithm with fuzzy c means (BOA-FCM) clustering. Journal of Medical Imaging and Health Informatics, 11(3), 661-666.

[Jansi & Subashini, 2014] Jansi, S., & Subashini, P. (2014, December). Modified FCM using genetic algorithm for segmentation of MRI brain images. In 2014 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-5). IEEE.

[Jardim et al., 2023] Jardim, S., António, J., & Mora, C. (2023). Image thresholding approaches for medical image segmentation-short literature review. Procedia Computer Science, 219, 1485-1492.

[Jasti et al., 2022] Jasti, V. D. P., et al. (2022). Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. Security and Communication Networks, 2022, 1918379.

[Jian et al., 2018] Jian, J., et al. (2018). Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images. Australasian Physical & Engineering Sciences in Medicine, 41, 393-401.

[Jin et al., 2021] Jin, Y., et al. (2021). 3D PBV-Net: an automated prostate MRI data segmentation method. Computers in Biology and Medicine, 128, 104160.

K

[Kahali et al., 2019] Kahali, S., Sing, J. K., & Saha, P. K. (2019). A new entropy-based approach for fuzzy c-means clustering and its application to brain MR image segmentation. Soft Computing, 23, 10407-10414.

[Kang et al., 2009] Kang, J., Min, L., Luan, Q., Li, X., & Liu, J. (2009). Novel modified fuzzy c-means algorithm with applications. Digital Signal Processing, 19(2), 309-319.

[Kang & Zhang, 2012] Kang, J., & Zhang, W. (2012). Combination of fuzzy C-means and particle swarm optimization for text document clustering. Advances in Electrical Engineering and Automation, 139, 247-252.

[Karaboga & Basturk, 2007] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal of Global Optimization, 39, 459-471.

[Karaboga & Ozturk, 2011] Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. Applied Soft Computing, 11(1), 652-657.

[Karthick et al., 2014] Karthick, S., Sathiyasekar, K., & Puraneeswari, A. (2014). A survey based on region based segmentation. International Journal of Engineering Trends and Technology, 7(3), 143-147.

[Katoch et al., 2021] Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications, 80, 8091-8126.

[Katarya & Verma, 2018] Katarya, R., & Verma, O. P. (2018). Recommender system with grey wolf optimizer and FCM. Neural Computing and Applications, 30, 1679-1687.

[Kennedy & Eberhart, 1995] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95-International Conference on Neural Networks, 4, 1942-1948.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. Science, 220(4598), 671-680.

[Kullback & Leibler, 1951] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics, 22(1), 79-86.

[Kumar & Kumari, 2018] Kumar, S., & Kumari, R. (2018). Artificial bee colony, firefly swarm optimization, and bat algorithms. In Advances in swarm intelligence for optimizing problems in computer science (pp. 145-182). Chapman and Hall/CRC.

[Kumar & al., 2024] Kumar, V. V., Divya, V., & Kumar, C. T. (2024, December). Hybrid fuzzy C-means and ant colony optimized (ACO) clustering approach for efficient routing in flying ad-hoc networks (FANETs). In 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1031-1037). IEEE.

L

[Li & Wen, 2015] Li, H., He, H., & Wen, Y. (2015). Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation. Optik, 126(24), 4817-4822.

[Liang et al., 2012] Liang, J., et al. (2012). Determining the number of clusters using information entropy for mixed data. Pattern Recognition, 45(6), 2251-2265.

[Lie & Bailey, 2014] Lie, Y. B., & Bailey, J. (2014). Generalized information theoretic cluster validity indices for soft clusterings. Proceedings of IEEE Symposium Series on Computational Intelligence, 24-31.

[Lingappa & al., 2018] Lingappa, H., Suresh, H., & Manvi, S. (2018). Medical image segmentation based on extreme learning machine algorithm in kernel fuzzy c-means using artificial bee colony method. Int. J. Intell. Eng. Syst, 11(6), 128-136.

[Liu et al., 2008] Liu, H. C., et al. (2008). Fuzzy C-mean clustering algorithms based on Picard iteration and particle swarm optimization. 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, 2, 616-619.

[Liu et al., 2021a] Liu, X., et al. (2021). Advances in deep learning-based medical image analysis. Health Data Science, 2021, 8786793.

[Liu et al., 2021b] Liu, Y., Jiang, Y., Hou, T., & Liu, F. (2021). A new robust fuzzy clustering validity index for imbalanced data sets. Information Sciences, 547, 579-591.

[Luenberger, 2016] Luenberger, D. G. (2016). Linear and Nonlinear Programming (4th ed.). Springer.

M

[Ma et al., 2014] Ma, J., et al. (2014). Fuzzy clustering with non-local information for image segmentation. International Journal of Machine Learning and Cybernetics, 5, 845-859.

[Maulik & Bandyopadhyay, 2003] Maulik, U., & Bandyopadhyay, S. (2003). Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. IEEE Transactions on geoscience and remote sensing, 41(5), 1075-1081.

[Mirjalili et al., 2014] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69, 46-61.

[Mittal et al., 2021] Mittal, H., et al. (2021). Gravitational search algorithm: a comprehensive analysis of recent variants. Multimedia Tools and Applications, 80, 7581-7608.

[Mohammdian-Khoshnoud et al., 2022] Mohammdian-Khoshnoud, M., Soltanian, A. R., Dehghan, A., & Farhadian, M. (2022). Optimization of fuzzy c-means (FCM) clustering in cytology image segmentation using the gray wolf algorithm. BMC Molecular and Cell Biology, 23(1), 9.

[Mohanapriya & Kalaavathi, 2019] Mohanapriya, N., & Kalaavathi, B. (2019). Adaptive Image Enhancement Using Hybrid Particle Swarm Optimization and Watershed Segmentation. Intelligent Automation & Soft Computing, 25(4), 713-724.

[Mokhtari & Debakla, 2018] Mokhtari, C., & Debakla, M. (2018, June 20-25). *Term classification based query expansion*. International Conference on Applied and Analysis Mathematic Modeling (ICAAMM2018), Istanbul, Turkey.

[Mokhtari et al., 2025] Mokhtari, C., Debakla, M., & Meftah, B. (2025). Fuzzy clustering optimization based artificial bee colony algorithm for brain magnetic resonance imaging image segmentation. *International Journal of Electrical and Computer Engineering (IJECE)*, 15 (5), 4916–4932.

[Mokhtari & Meftah, in press] Mokhtari, C., & Meftah, B. (in press). A new fuzzy clustering validity index based on Kullback-Leibler divergence. In *Proceedings of the International Conference on Artificial Intelligence: Theories and Applications (ICAITA 2025)*.

N

[Narayan et al., 2023] Narayan, V., et al. (2023). A comprehensive review of various approach for medical image segmentation and disease prediction. Wireless Personal Communications, 132(3), 1819-1848.

[Nesterov, 2018] Nesterov, Y. (2018). Lectures on convex optimization (Vol. 137). Berlin: Springer International Publishing.

[Ni, 2024] Ni, K. (2024). A clustering algorithm combining fuzzy C-means and artificial bee colony algorithm. International Journal of Innovative Computing, Information and Control, 20(1), 297-311.

P

[Pal et al., 2005] Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems, 13(4), 517-530.

[Parmar et al., 2018] Parmar, A., Katariya, R., & Patel, V. (2018). A review on random forest: An ensemble classifier. International Conference on Intelligent Data Communication Technologies and Internet of Things, 758-763.

[Pérez-Cruz, 2008] Pérez-Cruz, F. (2008). Kullback-Leibler divergence estimation of continuous distributions. IEEE International Symposium on Information Theory, 1666-1670.

[Pham et al., 2018] Pham, T. X., Siarry, P., & Oulhadj, H. (2018). Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. Applied Soft Computing, 65, 230-242.

[Poshitha et al.,2023] Poshitha, M., Ramaraj, K., Dilipkumar, S., Meena, R. D., Ambika, B., & Thilagaraj, M. (2023, December). Weighted fuzzy c means: A novel tumor segmentation approach in mr brain images. In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 574-581). IEEE.

[Prajapati et al., 2020] Prajapati, V. K., Jain, M., & Chouhan, L. (2020, February). Tabu search algorithm (TSA): A comprehensive survey. In 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE) (pp. 1-8). IEEE.

R

[Raghtate & Salankar, 2015] Raghtate, G. S., & Salankar, S. S. (2015, December). Modified fuzzy C means with optimized ant colony algorithm for image segmentation. In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 1283-1288). IEEE.

[Raj et al., 2024] Raj, J. R. F., et al. (2024). Brain tumor segmentation based on kernel fuzzy c-means and penguin search optimization algorithm. Signal, Image and Video Processing, 18(2), 1793-1802.

[Ray & Sing, 2024] Ray, M., & Sing, J. K. (2024, May). 3D brain MR image segmentation using a fuzzy entropy-based fuzzy clustering algorithm. In 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT) (pp. 327-331). IEEE.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention, 234-241.

[Rousseeuw, 1987] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65.

S

[Saha & Bandyopadhyay, 2009] Saha, S., & Bandyopadhyay, S. (2009). A new point symmetry based fuzzy genetic clustering technique for automatic evolution of clusters. Information Sciences, 179(19), 3230-3246.

[Sarkar et al., 2024] Sarkar, K., Mudi, R. K., & Pal, N. R. (2024). Weighted fuzzy C-means: Unsupervised feature selection to realize a target partition. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 32(08), 1111-1134.

[Saxena et al., 2019] Saxena, S., Jain, S., Tripathi, S., & Gupta, K. (2019, December). Comparative analysis of image segmentation techniques. In International Conference on Advanced Communication and Computational Technology (pp. 317-331). Singapore: Springer Nature Singapore.

[Schubert et al., 2017] Schubert, E., et al. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems, 42(3), 1-21.

[Seeram, 2015] Seeram, E. (2015). Computed tomography: Physical principles, clinical applications, and quality control. Elsevier.

[Sehmbi & Perlas, 2022] Sehmbi, H., & Perlas, A. (2022). Basics of ultrasound imaging. In Regional Nerve Blocks in Anesthesia and Pain Therapy (pp. 33-52). Springer.

[Semchedine & Moussaoui, 2018] Semchedine, M., & Moussaoui, A. (2018). An efficient particle swarm optimization for MRI fuzzy segmentation.

[Senthilkumaran & Vaithegi, 2016] Senthilkumaran, N., & Vaithegi, S. (2016). Image segmentation by using thresholding techniques for medical images. Computer Science & Engineering: An International Journal, 6(1), 1-13.

[Sharma et al., 2013] Sharma, P., Singh, G., & Kaur, A. (2013). Different techniques of edge detection in digital image processing. International Journal of Engineering Research and Applications, 3(3), 458-461.

[Shelokar et al., 2004] Shelokar, P. S., Jayaraman, V. K., & Kulkarni, B. D. (2004). An ant colony approach for clustering. Analytica chimica acta, 509(2), 187-195.

[Shrivastava & Bharti, 2020] Shrivastava, N., & Bharti, J. (2020). Automatic seeded region growing image segmentation for medical image segmentation: a brief review. International Journal of Image and Graphics, 20(03), 2050018.

[Song et al., 2018] Song, J., Cong, W., & Li, J. (2018). A robust fuzzy c-means clustering model with spatial constraint for brain magnetic resonance image segmentation. Journal of Medical Imaging and Health Informatics, 8(4), 811-816.

[Spall, 2005] Spall, J. C. (2005). Introduction to stochastic search and optimization: estimation, simulation, and control. John Wiley & Sons.

[Storn & Price, 1997] Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11(4), 341-359.

[Sujji et al., 2013] Sujji, G. E., Lakshmi, Y. V. S., & Jiji, G. W. (2013). MRI brain image segmentation based on thresholding. International Journal of Advanced Computer Research, 3(1), 97.

T

[Taha & Hanbury, 2015] Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging, 15(1), 29.

[Tan et al., 2023] Tan, C. H., et al. (2023). Hybrid fuzzy C-means using particle swarm optimization (PSO) and differential evolution (DE) for image segmentation. IEEE 19th International Conference on Automation Science and Engineering.

[Thomas & Kumar, 2024] Thomas, E., & Kumar, S. N. (2024). Fuzzy C means clustering coupled with firefly optimization algorithm for the segmentation of neurodisorder magnetic resonance images. Procedia Computer Science, 235, 1577-1589.

V

[van Erven & Harremoës, 2014] van Erven, T., & Harremoës, P. (2014). Rényi divergence and Kullback-Leibler divergence. IEEE Transactions on Information Theory, 60(7), 3797-3820.

[Veelaert & Teelen, 2009] Veelaert, P., & Teelen, K. (2009). Adaptive and optimal difference operators in image processing. Pattern recognition, 42(10), 2317-2326.

[Verger et al., 2021] Verger, A., et al. (2021). Single photon emission computed tomography/positron emission tomography molecular imaging for parkinsonism: A fast-developing field. Annals of Neurology, 90(5), 711-719.

[Viallon et al., 2015] Viallon, M., Cuvinciuc, V., Delattre, B., Merlini, L., Barnaure-Nachbar, I., Toso-Patel, S., ... & Haller, S. (2015). State-of-the-art MRI techniques in neuroradiology: principles, pitfalls, and clinical applications. Neuroradiology, 57(5), 441-467.

[Voudouris & al., 2010] Voudouris, C., Tsang, E. P., & Alsheddy, A. (2010). Guided local search. In Handbook of metaheuristics (pp. 321-361). Springer, Boston, MA.

W

[Wang & al, 2012] Wang, Q., Zhang, Q. P., & Zhou, W. (2012). Study on Remote Sensing Image Segmentation Based on ACA–FCM. Physics Procedia, 33, 1286-1291.

[Wang et al., 2017] Wang, X., et al. (2017). Noise-robust adaptive mean shift clustering. Pattern Recognition, 63, 235-246.

[Weng & Zhu, 2021] Weng, W., & Zhu, X. (2021). INet: Convolutional networks for biomedical image segmentation. IEEE Access, 9, 16591-16603.

[Winston, 1991] Winston, M. L. (1991). The biology of the honey bee. harvard university press.

X

[Xie, 1991] Xie, X. L. (1991). A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(8), 841-847.

Y

[Yang, 2009] Yang, X. S. (2009). Firefly algorithms for multimodal optimization. International Symposium on Stochastic Algorithms, 169-178.

[Yang, 2010] Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In Nature inspired cooperative strategies for optimization (NICSO 2010) (pp. 65-74). Berlin, Heidelberg: Springer Berlin Heidelberg.

[Yin, 2002] Yin, P. Y. (2002). Maximum entropy-based optimal threshold selection using deterministic reinforcement learning with controlled randomization. Signal Processing, 82(7), 993-1006.

[Yu et al., 2023] Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., ... & Gao, M. (2023). Techniques and challenges of image segmentation: A review. Electronics, 12(5), 1199.

Z

[Zahid et al., 1999] Zahid, N., Limouri, M., & Essaid, A. (1999). A new cluster-validity for fuzzy clustering. Pattern Recognition, 32(7), 1089-1097.

[Zaitoun & Aqel, 2015] Zaitoun, N. M., & Aqel, M. J. (2015). Survey on image segmentation techniques. Procedia Computer Science, 65, 797-806.

[Zhang et al., 2006] Zhang, X., Li, J., & Yu, H. (2006). On the use of similarity index for fuzzy clustering validity evaluation. Pattern Recognition Letters, 27(14), 1615-1621

[Zhang et al., 2017] Zhang, X., Sun, Y., Wang, G., Guo, Q., Zhang, C., & Chen, B. (2017). Improved fuzzy clustering algorithm with non-local information for image segmentation. Multimedia Tools and Applications, 76(6), 7869-7895.

[Zhang et al., 2021] Zhang, X., et al. (2021). Improved clustering algorithms for image segmentation based on non-local information and back projection. Information Sciences, 550, 129-144