T.C.

ISTANBUL AYDIN UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

SEGMENTATION OF LUNG COMPUTED TOMOGRAPHY IMAGES

MASTER'S THESIS

Ghaith TAYARA

Department of Engineering
Electrical and Electronics Engineering Program

JULY, 2022

T.C.
ISTANBUL AYDIN UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

SEGMENTATION OF LUNG COMPUTED TOMOGRAPHY IMAGES

MASTER'S THESIS

Ghaith TAYARA
(Y1913.300020)

Department of Engineering
Electrical and Electronics Engineering Program

Thesis Advisor: Asst. Prof. Dr. Necip Gökhan KASAPOĞLU

JULY, 2022

ONAY FORMU

DECLARATION

I hereby declare with the respect that the study “Segmentation of Lung Computed Tomography Images”, which I submitted as a Master thesis, is written without any assistance in violation of scientific ethics and traditions in all the processes from the project phase to the conclusion of the thesis and that the works I have benefited are from those shown in the References. (25/10/2022)

Ghaith TAYARA

FOREWORD

First, I would like to express my endless gratitude to God for being who I am right now and helping me to find patience, strength within myself to complete this thesis.

I would also like to thank my family not only for encouraging me to go abroad for a master’s degree but also for teaching me to chase my dreams and never give up.

I feel very fortunate to have Dr. Necip Gökhan KASAPOĞLU as my supervisor and want to express my appreciation for guiding me within the whole research process in a patient and effective manner.

Finally, I would like to acknowledge the important contribution of Istanbul Aydin University to my life, not only from an academic perspective but helping to meet great people that inspire, challenge, support and motivate me.

July, 2022 Ghaith TAYARA

SEGMENTATION OF LUNG COMPUTED TOMOGRAPHY IMAGES

ABSTRACT

The leading cause of cancer-related fatalities globally is lung cancer. Your airways, lung tissues, or blood that flow into and out of your lungs may all be affected by lung disorders. Early detection of lung disorders is essential, especially after the wake of the COVID19 epidemic. Therefore, the survival rate of patients is significantly influenced by early therapy. A system that can help radiologists identify CT scans is required to diagnose CT scans more quickly and to minimize human error. In this work, we compared three segmentation techniques: K-means clustering, Fuzzy C-means (FCM), as baseline methods, and Superpixel-based Fast Fuzzy C-means (SFFCM) as the proposed method, using synthetic images and lung CT scan images. Results showed that the proposed method (SFFCM) yields better segmentation results and has more robustness than the two baseline methods (K-means and FCM).
Keywords: Image Segmentation, Computed Tomography (CT) Images, Superpixel-based Fast Fuzzy C-means, K-means, Fuzzy C-means, Lung Cancer Detection

AKCİĞER BİLGİSAYARLI TOMOGRAFİ GÖRÜNTÜLERİNİN BÖLÜTLENMESİ

ÖZET

Dünyada kansere bağlı ölümlerin önde gelen nedeni akciğer kanseridir. Solunum yollarınız, akciğer dokularınız veya akciğerlerinize giren ve çıkan kanın tümü akciğer bozukluklarından etkilenebilir. Akciğer rahatsızlıklarının erken tespiti, özellikle COVID19 salgınının ardından çok önemlidir. Bu nedenle, hastaların hayatta kalma oranı, erken tedaviden önemli ölçüde etkilenir. BT taramalarını daha hızlı teşhis etmek ve insan hatasını en aza indirmek için radyologların BT taramalarını tanımlamasına yardımcı olabilecek bir sistem gereklidir. Bu çalışmada, üç segmentasyon tekniğini karşılaştırdık: K-ortalama kümeleme, temel yöntemler olarak Bulanık C-ortalamalar (FCM) ve önerilen yöntem olarak sentetik görüntüler ve akciğer BT kullanarak Süperpiksel tabanlı Hızlı Bulanık C-ortalamalar (SFFCM) görüntüleri tarayın. Sonuçlar, önerilen yöntemin (SFFCM) daha iyi segmentasyon sonuçları verdiğini ve iki temel yöntemden (K-ortalamalar ve FCM) daha fazla sağlamlığa sahip olduğunu göstermiştir.

Anahtar Kelimeler: Görüntü Bölütleme, Bilgisayar Tomografisi görüntüleri, Super piksel tabanlı Hızlı Bulanık C-Ortalar yöntemi, K-Ortalar, Akciğer Kanserinin belirlenmesi

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ALU	: Arithmetic Logic Unit
BSDS	: Berkeley segmentation dataset and benchmark
CAD	: Coronary Artery Disease
CNN	: Convolutional Neural Networks
CT	: Computed Tomography
FCM	: Fuzzy C-means
FCM_S	: Fuzzy C-means with special constraints
FCN	: Fully Convolutional Networks
FLICM	: Fuzzy Local Information C-means clustering algorithm
FRFCM	: fast and robust FCM algorithm
HMRF	: Hidden Markov random field
KWFLICM	: FLICM based on kernel metric and weighted fuzzy factor
MGR	: Morphological gradient reconstruction
MMGR	: Multiscale morphological gradient reconstruction
MRI	: Magnetic Resonance Imaging
NMR	: Nuclear Magnetic Resonance
NWFCM	: Neighborhood Weighted Fuzzy C-means clustering algorithm
OD	: Optical Density
RF	: Radiofrequency

SA : Segmentation accuracy

SE : Structural element

SFFCM : Superpixel-based fast FCM

SLIC : Simple linear iterative clustering

WT : Watershed

I. INTRODUCTION

The leading cause of cancer-related fatalities globally is lung cancer. (Shaziya, Shyamala, & Zaheer, 2018). Your airways, lung tissues, or blood flow into and out of your lungs may all be affected by lung disorders. In clinics, rapid or early detection of lung disorders is essential, particularly in the wake of the COVID19 epidemic. Therefore, the eventual survival rate of patients is significantly influenced by early therapy (Chunran, Yuanyuan, & Yi, 2018).

A system that can help radiologists identify CT scans is required in order to diagnose CT scans more quickly and to minimize human error (Silvana, Akbar, Gravina, & Firdaus, 2020). In such a system, automatic lung segmentation on CT images is often needed as a pre-processing step. Effective lung segmentation on CT images is essential for computer-aided diagnostic systems of lung disorders since it may influence the analysis that comes after.

To separate human body components from the backdrop and provide an initial categorization that clearly separates the right lung from the left lung, lung segmentation is performed. It is essential to segment the lungs since this is the first step in more quantitative lung analysis in CAD, such as measuring lung volume and analyzing lung texture to diagnose a particular condition.

A. Digital Image Processing

The definition of an image could be done using the two-dimensional function f(x,y); its amplitude at each given pair of coordinates (x,y) corresponds to the image's intensity or gray level at that location. A digital picture is made up of a limited number of pixels.

Processing digital images using a digital computer is referred to as the field of digital image processing (Rafael & Richard, 2008). The essential elements of an image processing system are shown in Figure 1 below. We will go through each element, beginning with the image sensors. There are two parts to it. The first is a physical instrument that can detect the energy emitted by the item (Rafael & Richard, 2008). The digitizer, the second component, is a device that turns the data gathered from the Sensor into a digital format.

The digitizer we previously stated and other hardware that executes other elementary functions, such as an Arithmetic Logic Unit, make up specialized image processing hardware (ALU). As an example, consider the use of ALU for image averaging during digitization to lower noise.

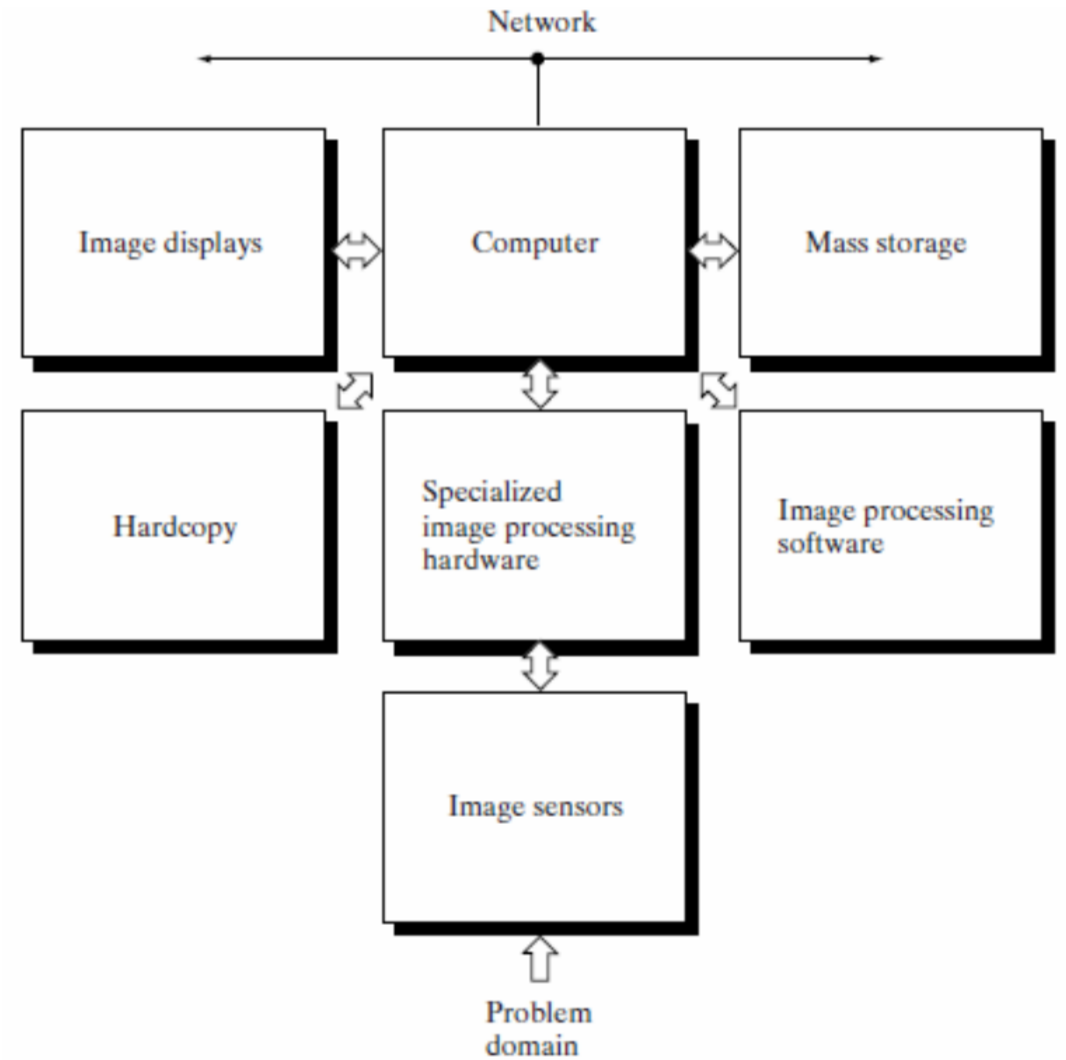

Figure 1: Components of general-purpose Image Processing System.

The computer here can be any computer, ranging from a normal PC to a supercomputer. Some applications may need a certain level of performance, and it differs from one application to another. Image processing software is a specialized modules that performs specific tasks (Rafael & Richard, 2008), a well known example of it is Matlab. Mass storage is the space needed for storing the digital image, for example: an image of size 1024 x 1024 pixels, with the intesity of each pixel is 8-bit, requires 1 megabyte of free storage.

Three types of digital storage can be used for image processing: short-term storage for usage during processing, online storage for quick recall, and archival storage, which has infrequent access (Rafael & Richard, 2008). Image displays, nowadays the most common are color TV monitors, which are an electronic devices with a screen used to display (as of television pictures or computer information).

Sometimes, the display card that is sold commercially as part of the computer system does not meet the requirements of particular image display applications [4]. Printers, film cameras, heat-sensing equipment, inject units, and digital equipment are examples of hardcopy devices. Any computer system has networking as a standard feature, and thankfully, nearly everyone now has access to the internet.

B. Medical Imaging

A form of energy is needed for the medical imaging of the human body. In radiology for example, the energy must penetrate the tissues in order to aquire the image. Since the visible light has a limited ability to penetrate the tissues, it is not used in the radiology department, it is used mostly in pathology (light microscopy),obstetrics (endoscopy), and dermatology (skin photography).

In terms of diagnostic radiology, we are referring to the use of electromagnetic radiation for medical imaging that is not in the visible light spectrum, such as x-rays in mammography and computed tomography (CT), radiofrequency (RF) in magnetic resonance imaging (MRI), and gamma rays in nuclear medicine. High-frequency soundwaves are the mechanical energy used in ultrasound imaging (Jerrold, J.Anthony, Edwin, & John, 2012). All medical imaging needs a form of energy to penetrate the tissues, except for nuclear medicine. The energy needs some sort of interaction like absorption or scattering so we can form an image, otherwise if the energy was only to pass through, then the energy that we detect would not be of any use for constructing an image.

In nuclear medical imaging, they inject the body with substances called Radionuclides, these include forms of the elements thallium, gallium, xenon, ionide, and thallium. Those are radioactive substances are injected or ingested and after that a radiation will be given off. A radiation detector picks up this radiation, physiological interactions will occur, and will give us information in the images.

C. Imaging Modalities

1. Radiography

Wilhelm Konrad Röntgen made the discovery of X-rays in 1895 when he was working with cathode tubes (Paul, 2009). By exposing a body to an X-ray phantom, an X-ray picture is a 2D projection of a 3D body on a film; these images are called projection on planar. In screen-film radiography, the optical density (OD) at a given spot on the film is (ideally) defined by the anatomy of the patient's x-ray attenuation along a straight line through the patient between the x-ray source and the spot on the detector.

Figure 2: Examples of X-ray Images.

2. Computed Tomography

The imaging technique known as X-ray computed tomography, or CT (Figure 3), creates cross-sectional pictures that show the body's X-ray attenuation characteristics. A computer reconstructs the organ under investigation in a sequence of cross sections or planes, then integrates X-ray images from multiple slices to rebuild 3D structures. Ct analyzes the attenuation of X-rays from various angles.

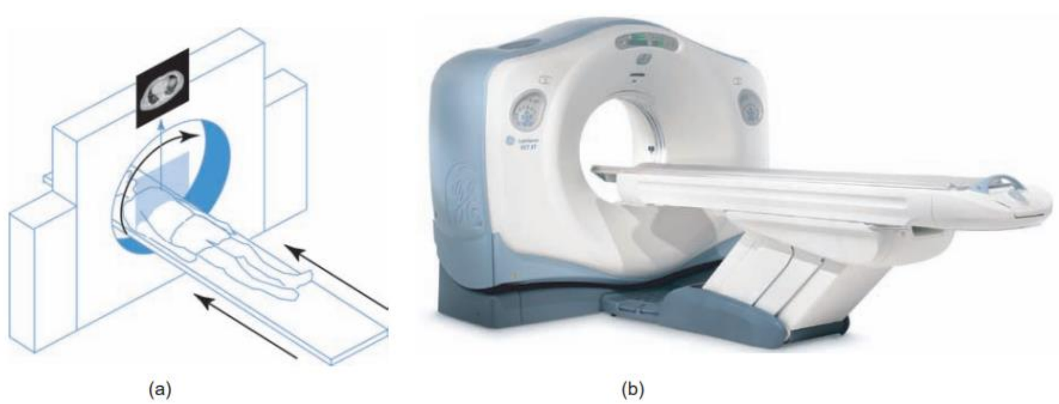

Figure 3: (a) Schematic representation, and (b) photograph of a CT scanner *(Paul, 2009)*.

3. Magnetic Resonance Imaging

Nuclear Magnetic Resonance is the basis of the modern medical imaging method known as MRI, which provides extensive information on the architecture of the soft tissues in humans (NMR). The main use for it in medical imaging is to show pathogenic or other physiological changes in live tissues. It may be used for both functional and anatomical imaging. It may be immediately recorded in any oblique plane and is really 3D.

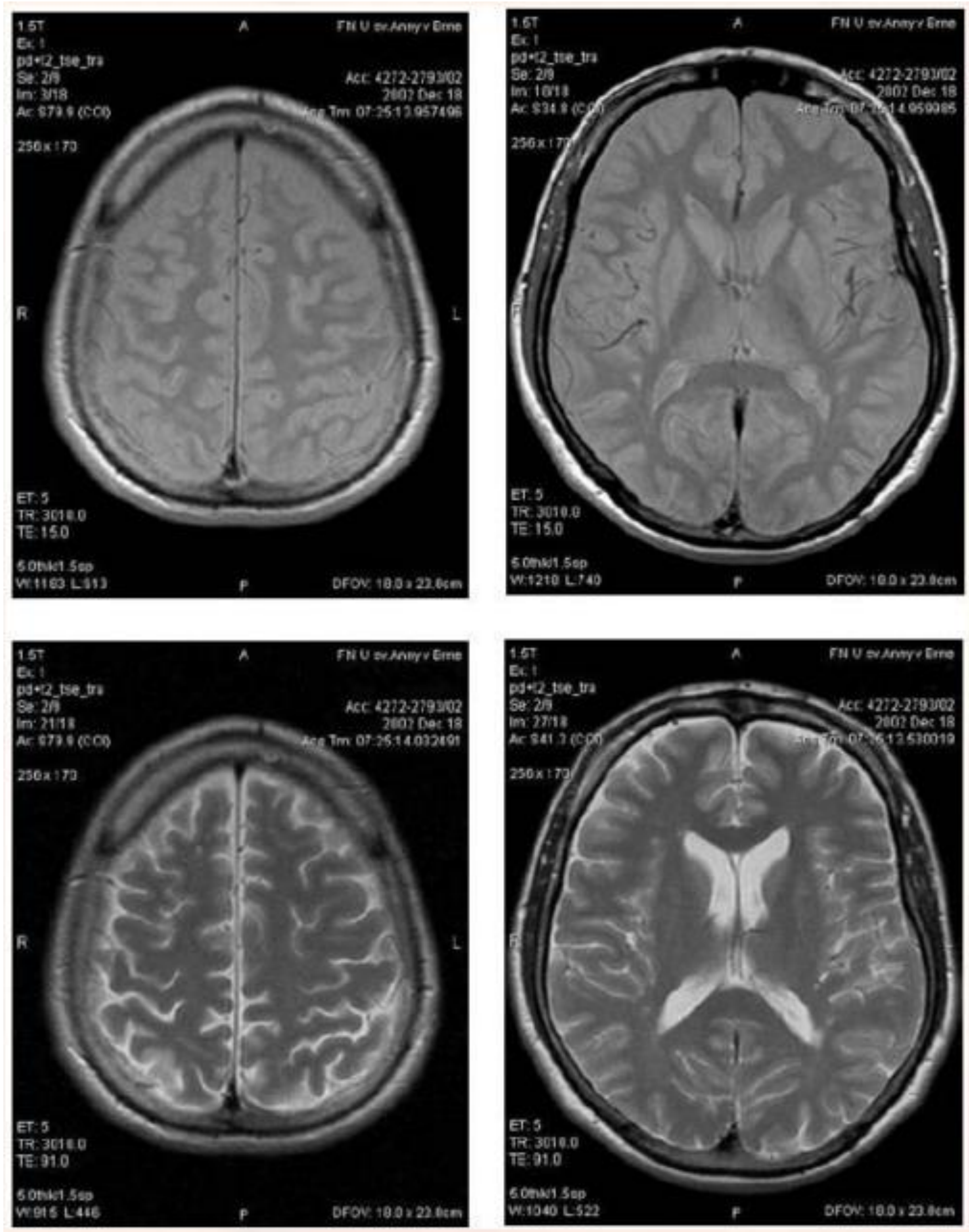

Figure 4: Examples of MRI images.

4. Ultrasound Imaging

Beyond the human auditory range's highest frequency of 20 kHz, ultrasound is a kind of acoustic wave. Medical sonography, often known as ultrasound, has a broad range of clinical uses as a main modality and as a contribute to other diagnostic techniques. It works by sending high frequency sound waves into the body, which are then received, processed, and shown parametrically as they bounce back from various organs and tissues.

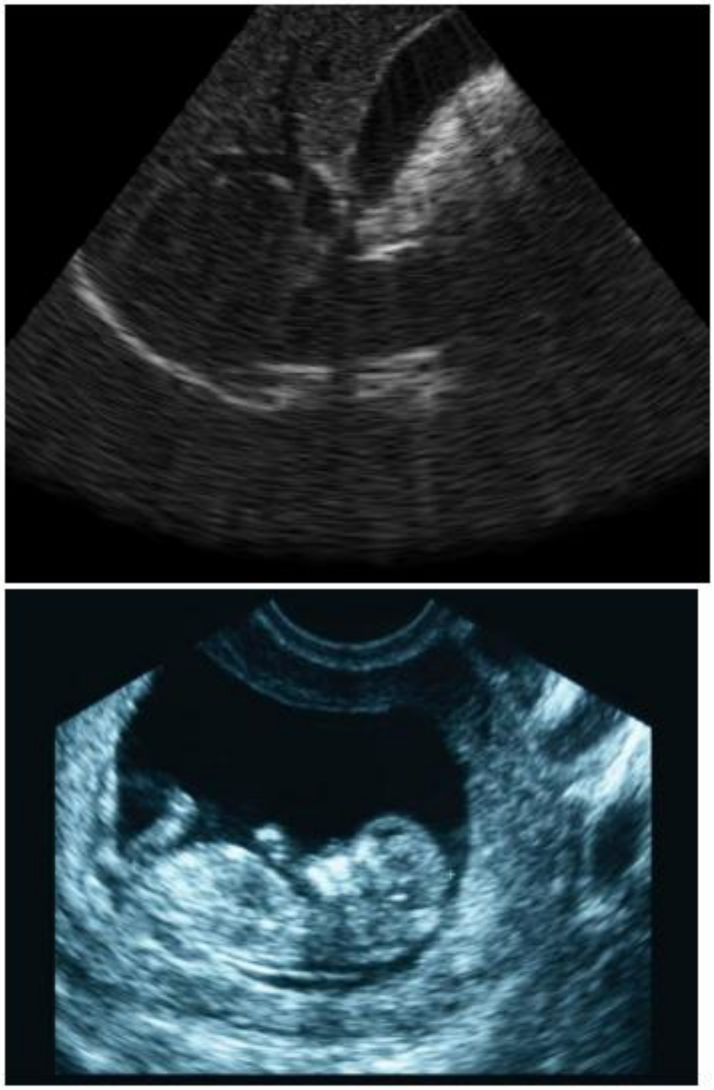

Figure 5: Example of Ultrasound imaging.

D. Image Segmentation

Image segmentation is the division of an image into separate, homogenous sections that don't overlap and have certain characteristics. When the application's items of interest have been separated, segmentation should come to an end. Attributes like gray level, color, texture, and form aid in the identification of areas, and groupings of regions with similar properties are created to represent certain meanings.

The process of segmenting a picture may be done using a variety of methods, including thresholding, clustering, region-based, edge-based, model-based, watershed approach, and many more. Combining approaches from different categories, such as those that combine edge detection and thresholding, may improve segmentation performance.

E. Lung Segmentation

Since the lung nodule is the early stage of the manifestations of lung cancer , the detection of it is very essential to clinical diagnosis of lung cancer , and the segmentation of the lung is the way to do that. The accuracy of lung segmentation is of great significance and importance because it can provide us with a precise information of the disease so we can follow up with the medical diagnosis and treatment. For the automated examination of a lung nodule that follows, lung segmentation is necessary for the estimation of lung volumes as well as the identification and measurement of lung anomalies within the lungs (Shaziya, Shyamala, & Zaheer, 2018).

Lung segmentation methods may be broadly categorized into distinct groups based on region, form, border, edge, threshold, and machine learning methods (Shaziya, Shyamala, & Zaheer, 2018).

F. Motivation

Despite the fact that the doctors are excellent at understanding the medical images to look for diseases, introducing and imroving medical image segmentation will assest the medical field and will help the doctors and the lab technicians interpreting medical images faster and easier. Moreover, automated systems are being improved to make it more reliable, efficient and faster.

Automated image segmentation is used for the extraction of the boundary features of the object, its also vital to understand the image context (Vagelis, 2009). However, its difficult to localize normal and abnormal regions in medical images, and it is a fundamental thing to do. Developing new segmentation techniques and applying them into the medical field will help and improve the healthcare sector, it will help physicians and lab technicians localize and identify tumers faster and easier, and this alone will help save lives, because the faster we can detect the tumor, the easier its treatment is going to be. In this work, i will be comparing three different segmentation techniques, Otsu’s method, K-means, and Superpixel fast fuzzy c-means, in order to find the optimum and best method to segment the CT lung images.

II. BASELINE METHODS

A. Segmentation Using K-means

The K-means approach is an unsupervised clustering method that divides the input data points into several groups according to how far apart they naturally are from one another (Suman & Avi). The k-means technique uses vector space data characteristics to detect grouping in the data. The K-means approach will then allocate each observation to one of the clusters once the user has chosen the K-means clustering's K number of clusters (Gareth, Daniela, Trevor, & Robert, 2021).

The results of utilizing various K values to accomplish K-means clustering on a synthetic example with 150 observations are shown in the picture below. The procedure of K-means clustering is simple , we will start by introducing some notations , Let C_1 , ... , C_K those are the indices of the observation for each cluster , and these sets has the following properties:

1- $C_1 \cup C_2 \ldots \cup C_K = \{1,\ldots,n\}$. Which means, at least one of the K clusters contains at least one of the observations.

2- $C_k \cap C_{k'} = \emptyset$ for all $k \neq k'$. Which means, there is no overlap among the clusters.

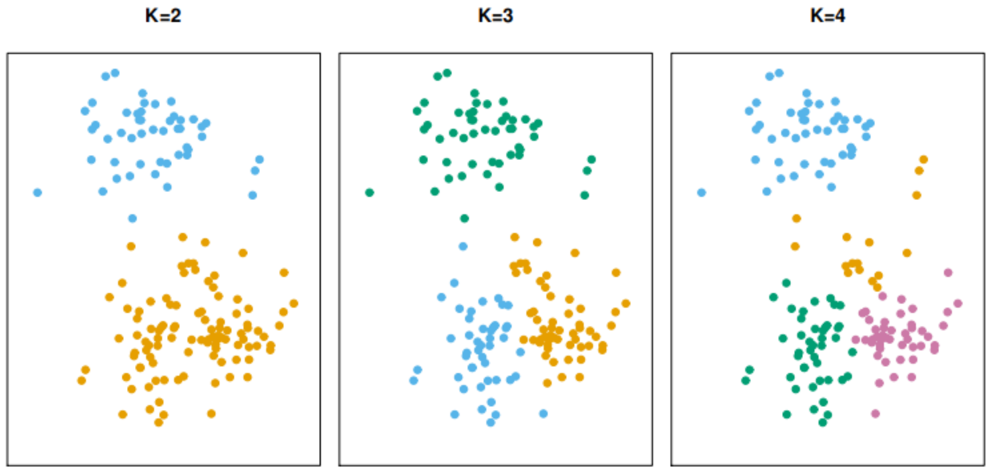

Figure 6: K-means clustering using different values of K on a simulated data set consisting of 150 observations in two dimensional space.

The primary principle of K-means clustering is that an effective clustering has as little within-cluster variance as feasible. (Gareth, Daniela, Trevor, & Robert, 2021).

$$\text{Minimize}(C_1,\ldots,C_k)\ \{ \textstyle\sum_{k=1}^{K} W(C_k)\ \} \tag{1}$$

The primary notion behind the K-means clustering is made clear by this formula, which divides the findings into K clusters, that the total inside cluster variance, added together over all K clusters, is as little as feasible (Gareth, Daniela, Trevor, & Robert, 2021).

B. Segmentation Using Fuzzy C-means

This segmentation technique is usually used in the analysis of data and information. Essentially, the algorith of K-means and Fuzzy C-means are somewhat similar as both algorithms uses of Euclidean distance for measurement. Instead, Fuzzy C-means algorithm works by classified membership to every one of the data points based on every cluster center (Iza, Noor, & Siti, 2020). The equation of the degree of membership is shown below.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(d_{ij}/d_{ik}\right)^{\left(\frac{2}{m}-1\right)}} \tag{2}$$

Where c denotes the cluster center, k denotes the iteration step, and μ_{ij} denotes the membership of the i^{th} data and j^{th} cluster center. Every piece of data is given membership in the algorithm based on its cluster. Firstly, we randomly select c which is the cluster center, then the fuzzy membership is computed using the previous equation. Finally V_j , which is the fuzzy centers is updated by performing the following equation , where V_j = 1,2,3,..,c :

$$V_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i)/(\sum_{i=1}^{n} ((\mu_{ij})^m) \quad (3)$$

One of the advantages of Fuzzy C-means that in contrast to hard segmentation, it preserves more information from the original picture. Using the Euclidean distance, fuzzy C-means can recognize the roughly shaped object (Iza, Noor, & Siti, 2020).

III. SUPERPIXEL-BASED FAST FUZYY C-MEANS CLUSTERING

A. Superpixels region-based segmentation

Superpixels segmentation increases the important information in comparison with signle pixels , thus making this approach more effective. It works by creating boundaries on the desired images without losing any quantitative information while segmenting (Iza, Noor, & Siti, 2020).

B. Superpixel-Based Fast Fuzzy C-means Clustering

1. Introduction

Fuzzy C-means is a very popular algorithm that often used to segment greysclae and color images, but it has two main problems. The repetitive computation of distances between clustering centers and pixels inside a local nearby window that results from the integration of local spatial information typically results in a high computational cost, and this is the first problem (Tao, ve diğerleri, 2019).

The Second problem is that the real local spatial structure of the images is being broken by the regular neighboring window, which leads to a poor segmentation. So combining Superpixel with Fuzzy C-means will result in a fast and more robust segmentation (Tao, ve diğerleri, 2019).

Many algorithms solve the problem of missing spatial information by integrating local spatial information into the objective function, but on the other hand increases the computational complexity and computational cost of the algorithm which may lead to increasing the run time for the program.

There are two advantages in combining Superpixel with Fuzzy C-means. The first one is that the capability of superpixel to accomplish segmentation of the images using their local spatial information. The second advantage of superpixel is its ability to decrease the number of dissimilar pixels in an image by substituting each pixel with the average value of the superpixel region (Tao, ve diğerleri, 2019). Unsupervised image segmentation and supervised image segmentation are the two groups of image segmentation algorithms.

Approaches like watershed transform, fuzzy entropy, clustering, and active contour model, are called unsupervised approach, and are good and widely known thanks to their uncomplicatedness without reliying on labels and training samples. Image segmentation can be accomplished using feature learning methods like fully convolutional networks (FCN) and convolutional neural networks (CNN), however these methods require a big number of training samples and labeled images. Moreover, the outcome of the segmentation has a coarse contour because FCN and CNN effectively accomplish image classification. In this section, we will primarily examine unsupervised image segmentation (Tao, ve diğerleri, 2019). Due to its suitability and use for both low-dimensional and high-dimensional data, clustering constitutes one type of significant and well-known unsupervised algorithm for segmenting grayscale and color image data.

Normally, we can divide clustering algorithms into three categories, attempting to minimize the objective function, graph theory, and decomposing a density function. Here, we will concentrate on clustering-based image segmentation by attempting to minimize the objective function, and FCM and k-means are clustering algorithms that uses this technique. Initial clustering centers or membership can affect k-means, due to its being a hard clustering algorithm. On the other hand, The soft clustering algorithm called FCM enhances k-means' weaknesses at the price of more repetitions. Nevertheless, each of FCM and k-means are afflicted by noise for the purpose of image segmentation, the local spatial information of pixels is lost.

Numerous enhanced clustering algorithms that directly reflect spatial information into their objective function have been offered in recent years as a solution to this issue (Tao, ve diğerleri, 2019), which can be categorized into two groups. The first one uses the effect of image segmentation by using neighborhood information of a center pixel through a window of constant size.

Some of examples of that is FCM_S which is Fuzzy C-means with special constraints, FCM_S1, FCM_S2, FLICM which is Fuzzy Local Information C-means clustering algorithm, NWFCM which is neighborhood weighted Fuzzy C-means clustering algorithm, FGFCM which is Fast generalized Fuzzy C-means algorithm, NDFCM, a Fuzzy C-means algorithm based on noise detection, Memon's algorithm, and KWFLICM, a FLISM based on kernel metric and weighted fuzzy factor in addition to the previously mentioned FLISM.

The benefit of such algorithms is that it is possible to compute the neighborhood information in advance, apart from of course FLICM and FCM_S, in order to decrease the computational difficulty. Still, a neighborhood window of a constant size and structure is incapable of fulfilling the demand for reliable image segmentation. Instead of using a window with a set size and structure, the second group makes use of adaptive neighborhood information, for instance, Liu’s algorithm, adaptive FLICM, and Bai’s algorithm.

The second set of algorithms acquires an increased resilience for noisy pictures and an enhanced segmentation effect than the first group because adaptive neighborhood information is stable with genuine image structuring information.

However, the neighborhood information of the matching membership, which is beneficial for enhancing classification result, is neglected by enhanced FCM algorithms, which only consider the neighborhood information of a picture. HMRF is a well known algorithm for solving the problem. Due to the fact that HMRF deems the prior state of the present membership, it acquires an improved outcome than FCM for image segmentation.

Additionally, Y. Zhang, G. Liu, and A. Wang in “Incorporating adaptive local information into fuzzy clustering for image segmentation” enhanced FCM algorithm by incorporating the distance between pixels and the distance between various regions collected by mean-shift into its objective function. It is clear that algorithms we talked about before enhance the effectiveness of image segmentation at the expense of raising computing difficulty and run time. So, the challenge is how can we effectively decrease computational complexity while maintaining local spatial information. T. Lei used membership filtering and morphological reconstruction to create the FCM algorithm (FRFCM), which is quick and reliable.

The approach is very speedy and produces an enhanced segmentation result compared to modern algorithms removing the need to repeatedly calculate the distance between pixels in the neighborhood window and clustering centers. Nonetheless, the FRFCM needs long run time for color image segmentation due to the reason that the computation of a color image's histogram is challenging.

We introduce a superpixel-based fast FCM (SFFCM) for color image segmentation to address the issue. The newly introduced algorithm can segment colored images with a high degree of precision and a very low computing cost. The following two contributions are provided:

1) In order to create superpixel images with precise bounds, we introduce a multiscale morphological gradient reconstruction (MMGR) method. This operation is useful for incorporating adaptive surrounding information and decreasing the number of diverse pixels in a color image.

2) We recommend a simple color histogram computational technique based on a superpixel picture acquired by MMGR that can be utilized to accomplish a quick FCM algorithm for color image segmentation.

2. Methodology

Mean-shift, simple linear iterative clustering (SLIC), and WT are examples of superpixel technologies that are often regarded as presegmentation techniques for enhancing segmentation outcomes produced by clustering algorithms. This is because a superpixel picture can give greater local spatial information than a nearby window of set size and shape. Mean-shift and WT generate more irregular superpixel regions than SLIC does, which is preferable to the hexagonal zones that SLIC produces. Because WT is noise-sensitive and causes significant oversegmentation in real applications, mean-shift is much often used. Mean-shift may provide greater superpixel results, although it is susceptible to parameter values, such as the spatial bandwidth, range bandwidth, and minimum size of final output areas (h_s, h_r, and h_k).

Furthermore, mean-shift has a larger computational cost than WT. We must thus create a quick superpixel method capable of providing improved presegmentation outcomes while running faster than mean-shift. WT seems to have a very low computing cost since it solely relies on the area minima of gradient pictures to produce presegmentation. To create superpixel pictures in this study, we use an unique WT built on the MMGR algorithm (MMGR-WT). When compared to mean-shift, the MMGR-WT can deliver results for presegmentation that are more suitable.

Additionally, it is unaffected by parameters. We calculate the superpixel image histogram using the superpixel image acquired by MMGR-WT in order to create a fast FCM method. Since there are fewer colors in superpixel pictures than there are in the original color image, computing the histogram of superpixel pictures is simple. Lastly, in order to perform quick color picture segmentation, the histogram is taken into account as a variable of the objective function. Figure 7 displays the conceptual structure of the algorithm we suggested.

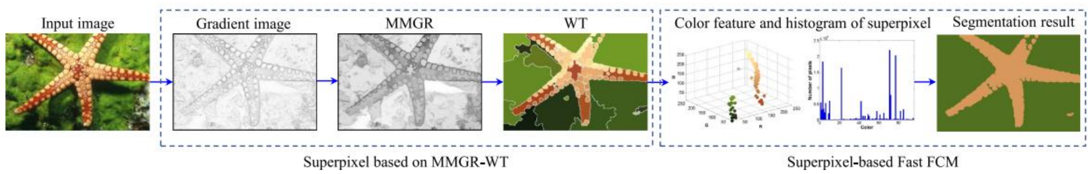

Figure 7: Structure of the introduced algorithm *(Tao, ve diğerleri, 2019)*.

a. Superpixel-Based on MMGR-WT

The quick method WT computes the local minima of a gradient picture and looks for the watershed line connecting nearby local minima. Due to its sensitivity to noise, the method easily leads to an oversegmentation. Numerous techniques have been suggested as solutions to the issue by changing the gradient picture of the original image.

A straightforward and effective approach for combating oversegmentation is morphological gradient reconstruction (MGR) (Luc, 1993), which is able to maintain object contour information while reducing noise and pointless gradient details. Firstly, the fundamental description of morphological reconstruction is given:

$$\begin{cases} R_f^{\varepsilon}(g) = \varepsilon_f^{(i)}(g) \\ R_f^{\delta}(g) = \delta_f^{(i)}(g) \end{cases} \tag{4}$$

Where f is the original picture, also known as the mask image, g is the marker image, ε is the erosion operation, and δ is the dilation operation, Rε and Rδ denote morphological erosion and dilation reconstruction, respectively.

Dilation reconstruction needs $g \leq f$, $\varepsilon^{(1)}{}_f\,(g) = \varepsilon(g) \vee f$, $\varepsilon\,^{(i)}{}_g\,(f) = \varepsilon(\varepsilon^{(i-1)}(g)) \vee f$, $\delta^{(1)}{}_f\,(g) = \delta(g) \wedge f$, and $\delta^{(i)}{}_g\,(f) = \delta(\delta^{(i-1)}(g)) \wedge f$. but erosion reconstruction also needs $g \geq f$. The pointwise maximum and minimum are represented by the symbols ∧ and ∨, respectively.

As a pair of dual operators, morphological erosion and dilation are always found in pairs, such as morphological opening and closing operators. Because they offer a better capacity for feature extraction or noise reduction, morphological opening and closure are more common than erosion and dilation. As a result, the following is the definition of the morphological opening reconstruction represented by R^O and closure reconstructions defined by R^C:

$$\begin{cases} R^{O}(g) = R^{\delta}(R^{\varepsilon}) \\ R^{C}(g) = R^{\varepsilon}(R^{\delta}) \end{cases} \tag{5}$$

Where the marker photo is Typically, g is thought of as $g = \varepsilon_B$ (f) in R^{δ} or $g = \delta_B(f)$ in R^{ε}. A structural element (SE) is B. To lessen oversegmentation, both R^O and R^C may eliminate area minima in a gradient picture (Tao, ve diğerleri, 2019).

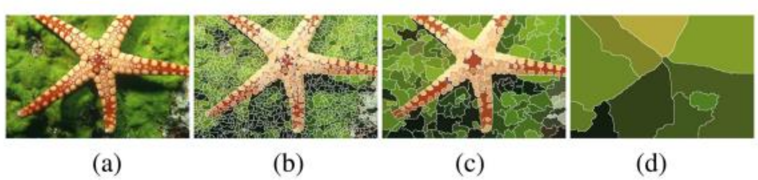

Figure 8: Segmenting a watershed using MGR and several SEs. (a) The original photo "12003" (481 x 321) (b) r = 1. (c) r = 3. (d) r = 10 *(Tao, ve diğerleri, 2019)*.

Figure 8 illustrates how we utilize R^C to reduce oversegmentation. The SE is shown as a disk in Figure 8, where r denotes the SE's radius. Figure 8 demonstrates that when the value of r is increased, the number of segmentation areas rapidly decreases. A big SE, on the other hand, is readily followed by undersegmentation, while a small SE is easily followed by oversegmentation.

As a result, it is challenging to produce a superpixel picture using MGR that has both fewer areas and a precise contour. An appropriate SE is needed to maintain the number of regions in the superpixel picture and the contour accuracy, however selecting a appropriate SE for various images may be challenging.

To address the issue, we attempt to rebuild a gradient picture using several SEs, fuse the resulting gradient images, and eliminate the dependence of the segmentation result on the SEs. As a result, we suggest an MMGR operation with the symbol R^{MC} with the following definition.

$$R_f^{MC}(g, r_1, r_2) = \vee \{R_f^C(\mathrm{g})B_{r1}, R_f^C(\mathrm{g})B_{r1+1}, \ldots, R_f^C(\mathrm{g})B_{r2}\} \quad (6)$$

Where r_1 and r_2 stand in for the minimum and maximum values of r, respectively, and where $r_1 \leq r \leq r_2$, r_1 , $r_2 \in N^+$, $g \leq f$. As we can see, R^{MC} uses many reconstructed pictures to create a gradient image by using multiscale SEs. An excellent gradient picture is created by finding the pointwise maximum of these rebuilt gradient images. This gradient image eliminates the majority of unnecessary local minima while retaining crucial edge information. Two parameters, r_1 and r_2, which regulate the size of the minimum area and the maximum region, respectively, are part of the proposed MMGR. The segmentation results will include a lot of little areas if r_1 is too small, but if r_1 is too big, the boundary accuracy will be poor. Figure 9 offers an example.

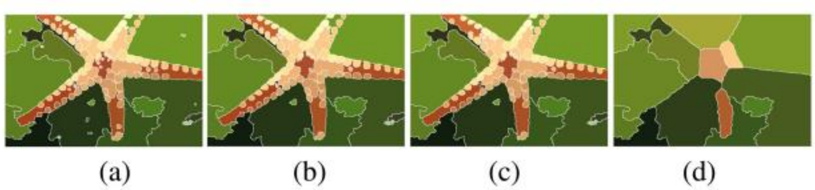

Figure 9: Results of segmentation using MMGR-WT with various r_1 values, when r_2 = 10. (a) r_1 = 1. (b) r_1 = 3. (c) r_1 = 5, (d) r_1 = 8 *(Tao, ve diğerleri, 2019)*.

The superpixel output may be observed to have a high contour accuracy but some tiny areas. When $r_1 = 1$, the superpixel outcome has a high contour accuracy and does not include tiny areas. However, while $r_1 = 2$ or $r_1 = 3$, the superpixel outcome has a distinctly poor contour precision. So, in this case, we choose $1 \leq r_1 \leq 3$. Because r_2 regulates the maximum region's size, a bigger r_2 value produces a better superpixel picture, as can be seen in Figure 10.

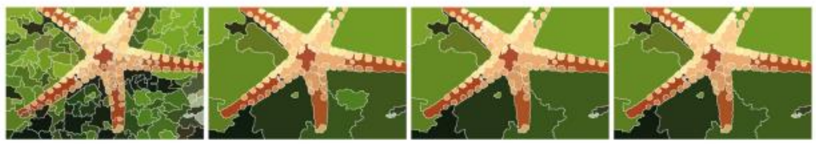

Figure 10: Watershed segmentation based on MMGR-WT with different sized SEs. (a) $r_1 = 2$, $r_2 = 3$. (b) $r_1 = 2$, $r_2 = 7$. (c) $r_1 = 2$, $r_2 = 11$, (d) $r_1 = 2$, $r_2 = 20$ *(Tao, ve diğerleri, 2019).*

Once the value of r_2 exceeds a threshold, as in Figure 10 when the threshold is 11, the superpixel picture remains unaltered. By raising the value of r_2, it is evident that the superpixel picture is convergent. The convergent outcome is also ideal since it has fewer areas while still producing precise contour. As a result, when r_2 is greater than a threshold, the MMGR is not affected by changes in r_2. Comparing the number of superpixel areas for WT-MGR and WT-MMGR, respectively, is shown in Table 1.

	MGR			MMGR		
Parameters	$r = 1$	$r = 3$	$r = 10$	$r_1 = 2, r_2 = 7$	$r_1 = 2, r_2 = 11$	$r_1 = 2, r_2 = 20$
Number	1210	263	10	264	95	95

Table1: Comparison of WT's superpixel areas based on MGR and MMGR, respectively (Tao, ve diğerleri, 2019).

Table 1 demonstrates that r_2 could be a variation. However, it is challenging to establish various r_2 values for each picture. If we choose a low error threshold, indicated by η rather than r_2, then r_2 is not necessary for MMGR in real situations, for example (Tao, ve diğerleri, 2019),

$$\max\{R_f^{MC}(g, r_1, r_2) - R_f^{MC}(g, r_1, r_2 + 1) \leq \eta\} \tag{7}$$

Since r_2 should have a distinct value for each picture in a data set, a fixed value of η may be used for all the images in the data set instead of r_2 in equation (7). The error will be big if η is too big, but r_2 will be modest. The error will be minor but r_2 will be big if η is too small, which will place a heavy computational strain on MMGR. Therefore, picking a suitable η for a data set is crucial.

As shown in Table 2, when we apply MMGR to 10 pictures from the Berkeley segmentation dataset and benchmark (BSDS), we may get various r_2 values based on a constant value of η.

Images	$\eta = 10^{-2}$	$\eta = 10^{-3}$	$\eta = 10^{-4}$	$\eta = 10^{-5}$
"2092"	12	17	26	26
"3096"	10	10	10	10
"8023"	10	10	14	14
"8049"	14	19	22	22
"8143"	7	10	10	10
"12003"	12	18	18	18
"12074"	10	18	24	24
"12084"	14	15	15	15
"14037"	10	14	17	17
"15004"	14	18	18	18

Table 2 : R2 values for 10 photos from the BSDS for various η values (Tao, ve diğerleri, 2019).

Table 2 demonstrates that as η is decreased, the values of r_2 will increase. However, when η is less than or equal to 10^{-4}, r_2 will remain unaltered. Therefore, in this study, we set = 10^{-4}.

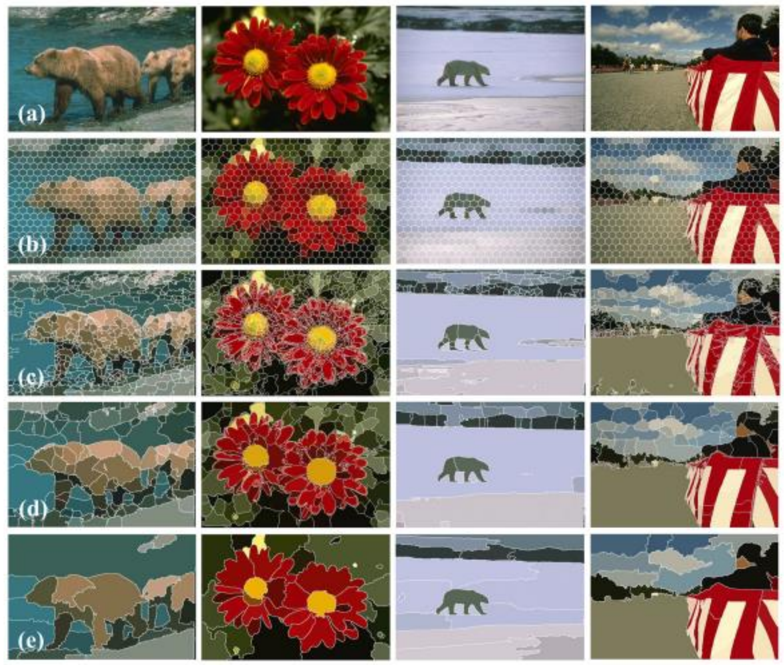

Figure 11: Several techniques for producing superpixel pictures. (a) Original photos. (b) SLIC-produced superpixel pictures with s_k = 500, s_m = 50, and s_s = 1. (c) Superpixel pictures created using mean-shift1 with the parameters h_s = 7, h_r = 7, and h_k = 30. (d) Superpixel pictures produced via mean-shift2 with h_s, h_r, and h_k set to 15, 15, and 50, respectively. (e) Superpixel pictures from the MMGR-WT algorithm (r_1 = 2). *(Tao, ve diğerleri, 2019)*.

Figure 11 exhibits superpixel pictures produced by SLIC, mean-shift, and MMGRWT, accordingly, to indicate the efficiency of the MMGR. In which s_k is the required number of superpixels, s_m is the weighting factor among color and spatial disparities, and s_s is the threshold being used for combining regions.

As observed in Figure 11, the superpixel pictures produced by SLIC have many regions of the same shape and size, but those produced by mean-shift and MMGR-WT contain many areas of various sizes and forms.

It is obvious that the latter two algorithms provide a greater visual impression when genuine photos are required. The suggested MMGR has a shorter runtime than SLIC, mean-shift, mean-shift1, and mean-shift2, as indicated in Table 3, where SLIC relates to Figure 11(b), mean-shift1 relates to Figure 11(c), mean-shift2 relates to Figure 11(d), and MMGR-WT relates to Figure 11(e). MMGR is a better choice than SLIC and mean-shift for our job requirement since our goal is to provide a quick FCM method for color picture segmentation.

Algorithms	"100075"	"124084"	"100007"	"145086"	Average
SLIC	3.86	4.07	3.89	3.88	3.93
mean-shift1	1.02	1.22	0.94	1.20	1.10
mean-shift2	2.66	1.22	2.67	2.86	2.79
MMGR-WT	**0.32**	**0.32**	**0.31**	**0.36**	**0.33**

The best values are in bold.

Table 3: Comparison of superpixel image generation time (in seconds) for various techniques.

b. Superpixel-Based Fast Fuzzy c-Means

The combining of MMGR-WT and FCM may enhance picture segmentation results since MMGR-WT relies on the local feature of an image whereas FCM relies on the global feature. In this part, we provide an SFFCM method that modifies the FCM's objective function to take into account the adaptation local spatial information. EnFCM is useful and successful for generating rapid picture segmentation since a gray picture has just 256 gray levels, that is often significantly less than the number of pixels in a picture, but a color image contains way more than 256 distinct colors.

Typically, quantization is used to minimize the amount of colors in a picture. Every channel of a color picture is subjected to a clustering technique in order to produce an image with less color levels than beforehand. Nevertheless, standard color quantization merely decreases the number of distinct colors, and the quantized picture's color distribution remains comparable with that of the main image since local spatial information is omitted.

A superpixel picture is preferable than photos quantized using clustering techniques because it retains the spatial information of the picture and minimizes the number of distinct colors. We used the clustering approach presented in (Doğan & Lale, 2002) and the suggested MMGR-WT to quantize a color picture, and then calculated the histogram of such quantized picture as demonstrated in Figure 12.

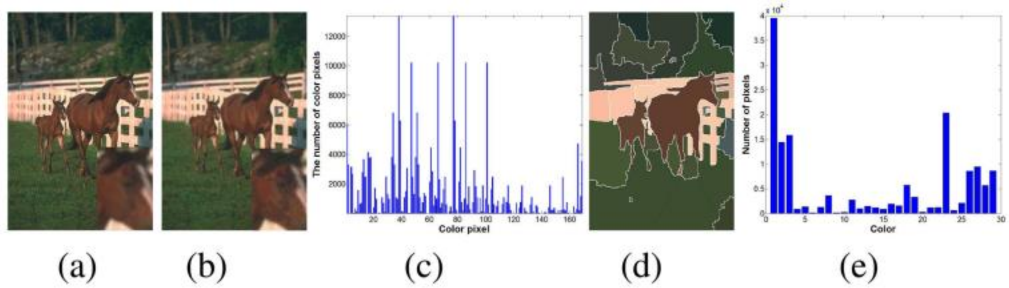

Figure 12: An picture in color and its related histogram are quantized. (a) The original photo. (b) Quantization of color using the approach suggested in *(Doğan & Lale, 2002)* with c = 10. (c) Figure 3.1's histogram (b). Superpixel picture created using MMGR-WT (r1 = 2) in (d). Figure 3.1's histogram (c) *(Tao, ve diğerleri, 2019)*.

Additionally,we can see in figure 13, which is the color distribution in figure 12, that MMGR-WT is much suitable in comparison with the clustering algorithm offered in (Doğan & Lale, 2002) for subsequent image segmentation.

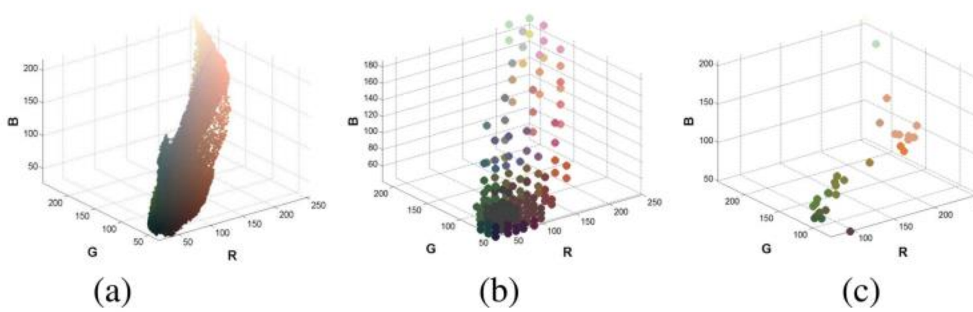

Figure 13: Distribution of colors in various color pictures. (a) The color distribution in Figure 12(a). (b)The color distribution of Figure 12(b). (c) The color distribution of Figure 12(d) *(Tao, ve diğerleri, 2019)*.

It is evident that the histograms in Figure 12(b) and (d) are less complicated since there are less colors present in the quantized photos. We can simply expand EnFCM to color image segmentation, as shown in Figure 12(c) and (e). Figure 12(e) contains even less color levels than Figure 12(c).

Additionally, it is obvious that Figure 13(C) color 's ranges vary from Figure 13(a) and (b), and the former is useful for a later categorization of pixels. Our suggested SFFCM objectives function for color picture segmentation is as follows, based on the superpixel photo produced by MMGR-WT:

$$J_m = \sum_{l=1}^{q} \sum_{k=1}^{c} S_1 u_{k1}^m \left\| \left(\frac{1}{S_1} \sum_{p \in R_1} x_p \right) - v_k \right\| \wedge 2 \tag{8}$$

Where S_l is the number of pixels in the lth region R_l, l, q $\in N^+$, and x_p is the color pixel inside the l^{th} region of the superpixel picture created by MMGR-WT. l is the color level. Comparison to the previous objective function in FCM, the updated objective function just adds histogram information.

The number of color levels is equal to the number of regions in the superpixel picture since every color pixel inside the original picture is substituted by the average value of color pixels inside the related area of the superpixel picture. Therefore, l << N effectively reduces the computation time and complexity. The preceding optimization issue may be transformed into an unrestrained optimization problem that reduces the subsequent objective function using the Lagrange multiplier approach.

$$J_m = \sum_{l=1}^{q} \sum_{k=1}^{c} S_1 u_{k1}^m \left\| \left(\frac{1}{S_1} \sum_{p \in R_1} x_p \right) - v_k \right\|^2 - \lambda \left(\sum_{k=1}^{c} u_{kl} - 1 \right) \tag{9}$$

When the Lagrange multiplier λ is used. With regard to u_{kl} and v_k, we construct the partial differential equation of J_m, accordingly.

$$\frac{\partial \widetilde{J_m}}{\partial u_{kl}} = \sum_{l=1}^{q} \sum_{k=1}^{c} \frac{\partial S_1 u_{kl}^m \left\| \left(\frac{1}{S_1} \sum_{p \in R_1} x_p \right) - v_k \right\|^2}{\partial u_{kl}} - \lambda$$

$$= \sum_{l=1}^{q}\sum_{k=1}^{c} mS_1 u_{kl}^{m-1} \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\|^2 - \lambda = 0 \tag{10}$$

$$\frac{\partial \widetilde{J_m}}{\partial u_{kl}} = \sum_{l=1}^{q}\sum_{k=1}^{c} \frac{\partial S_1 u_{kl}^{m} \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\|^2}{\partial v_k}$$

$$= \sum_{l=1}^{q}\sum_{k=1}^{c} S_1 u_{kl}^{m} \frac{\partial \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\|^2}{\partial v_k}$$

$$= \sum_{l=1}^{q} S_1 u_{kl}^{m} \frac{\partial \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\|^2}{\partial v_k}$$

$$= -2 \sum_{l=1}^{q} S_1 u_{kl}^{m} \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\| = 0 \tag{11}$$

The appropriate answers for u_{kl} and v_k are derived by combining (10) and (11) as follows.

$$v_k = \frac{\sum_{l=1}^{q} u_{kl}^{m} \sum_{p\in R_1} x_p}{\sum_{l=1}^{q} S_1 u_{kl}^{m}} \tag{12}$$

$$u_{kl} = \frac{\left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_k \right\|^{-2/(m-1)}}{\sum_{j=1}^{c} \left\| \left(\frac{1}{S_1} \sum_{p\in R_1} x_p \right) - v_j \right\|^{-2/(m-1)}} \tag{13}$$

The suggested SFFCM algorithm may be summed up as follows relying on (9)–(13).

Setting values for c, m, r1,η,η', and in Step 1, where η' is the SFFCM's convergence requirement. Using (6)–(7), create a superpixel picture in step 2 and then calculate its histogram.

1) Use the Sobel operations to calculate the gradient picture.

2) Put MMGR into practice using (9)–(10) and η.

3) To acquire the superpixel picture, use WT.

In Step three, Use the superpixel picture to randomly initialize the membership partition matrix $U^{(O)}$.

Put the loop counter to b = 0 in step four.

Modify the clustering centers according to (12) in step five.

Modify the membership 's partition matrix in step six by using (13).

In Step (7), If max{$U^{(b)}$ max$U^{(b+1)}$}< η', stop; if not, set b = b + 1 and go to Step five (Tao, ve diğerleri, 2019). We utilized the offered method SFFCM to figure 12(a) trailing the earlier steps. After that the segmentation outcomes are shown in figure 14.

(a) (b)

Figure 14: Results of segmentation on Fig. 6 (a). (a) The quantized picture segmentation result using FCM. (a) The results of segmentation utilizing the suggested SFFCM *(Tao, ve diğerleri, 2019)*.

We can observe that the suggested SFFCM outperforms the conventional method in terms of segmentation results. We draw the following conclusions about the suggested SFFCM based on the preceding analysis:

1) Superpixel and the color histogram effectively minimize the amount of various colors, making SFFCM quite quick at segmenting color images.

2) SFFCM is resistant to parameter changes due to the convergent superpixel picture produced by MMGR-WT.

3) SFFCM achieves an outstanding output for color picture segmentation since the objective function includes both adaptive local spatial information and a global color feature.

IV. RESULTS

After introducing three segmentation techniques (two baseline methods and one proposed method), which are K-means clustering, Fuzzy c-means, and Superpixel fast fuzzy c-means, we will start to compare between these three algorithms from different angles. First, we will work with synthetic images to measure the segmentation accuracy and the robustness of each technique. Gaussian and Salt and Pepper noise will be added to the image in different magnitudes: 5,10,15, and 20% percent.

A. Working on Synthetic Images

We've created a synthetic image to help us test the three algorithms, measure the optimal segmentation accuracy, and the check the robustness of each algorithm when subjected to a noisy images. The synthetic image consists of 5 regions, each one with a different grey level, as we can see in the figure below:

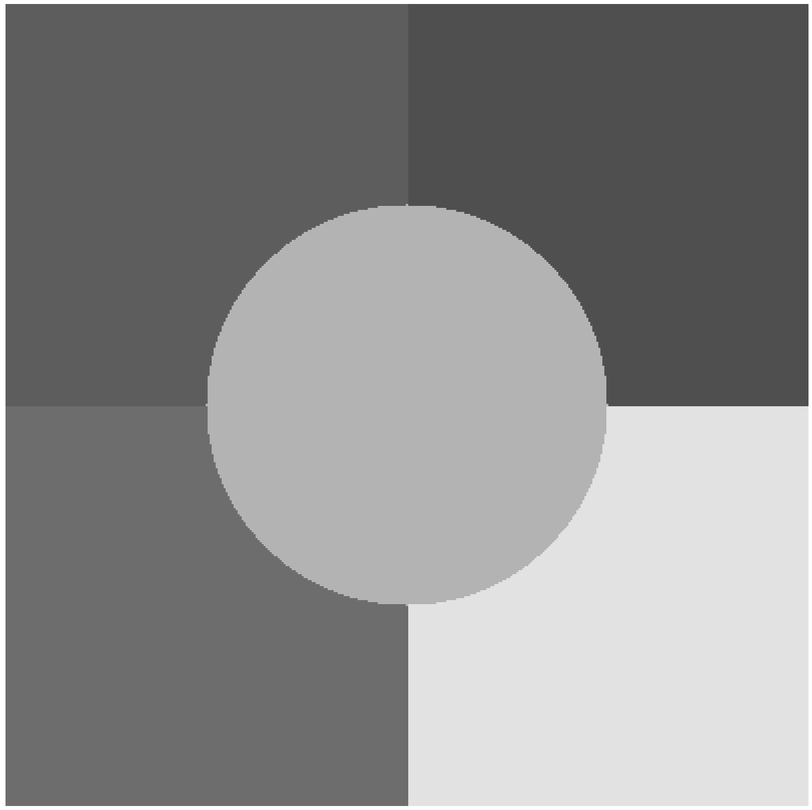

Figure 15: Synthetic image with 5 different grey levels.

B. Accuracy measures

To be able to check the performance for each algorithm and see how well it performs, there must be some mathematical measurements in order to help us compare between the three algorithms. We will adopt two performance indices, the optimal segmentation accuracy (SA), and the degree to which the segmented picture and the ground truth are identical; the quantitative score (S). Both are defined as:

$$S = \sum_{k=1}^{c} \frac{A_k \cap C_k}{A_k \cup C_k} \tag{14}$$

$$SA = \sum_{k=1}^{c} \frac{A_k \cap C_k}{\sum_{j=1}^{c} C_j} \tag{15}$$

in which C_k is the set of pixels corresponding to the class in the Ground Truth and A_k is the set of pixels corresponding to the k^{th} class that the algorithm has detected.

C. Testing on Synthetic Images

We will test each algorithm with two kinds of noisy images, Gaussian, and salt and pepper noise, each with four different levels of noise saturation: 5,10,15, and 20%.

1. Testing K-means Algorithm

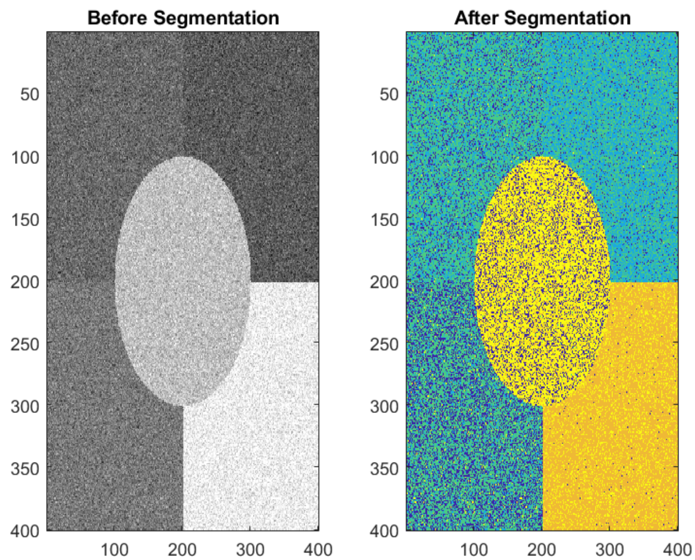

Figure 16: Synthetic image with 5% Gaussian noise before and after segmentation using k_means.

S= 35.6030%, SA= 48.2976%.

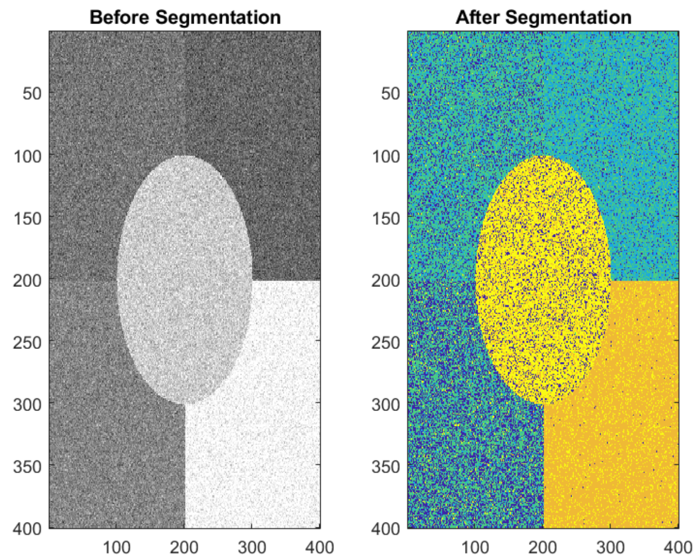

Figure 17: Synthetic image with 10% Gaussian noise before and after segmentation using k_means.

S= 35.5600%, SA= 47.8971%.

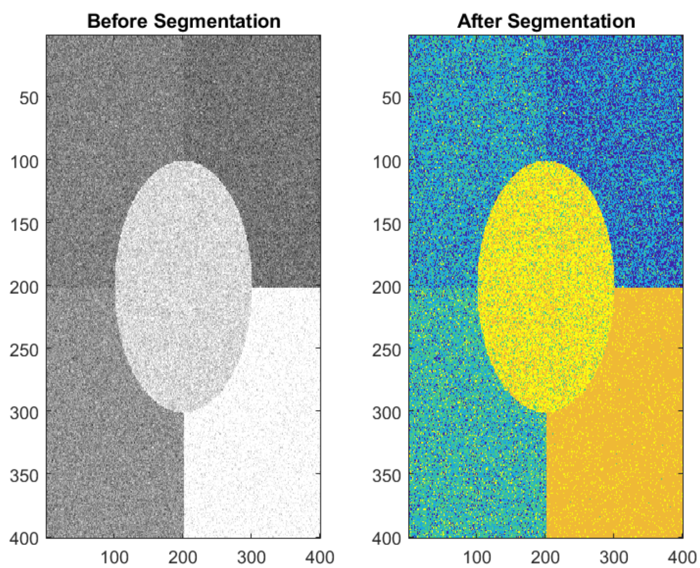

Figure 18: Figure 4.4: Synthetic image with 15% Gaussian noise before and after segmentation using k_means.

S= 34.6780%, SA= 47.6546%.

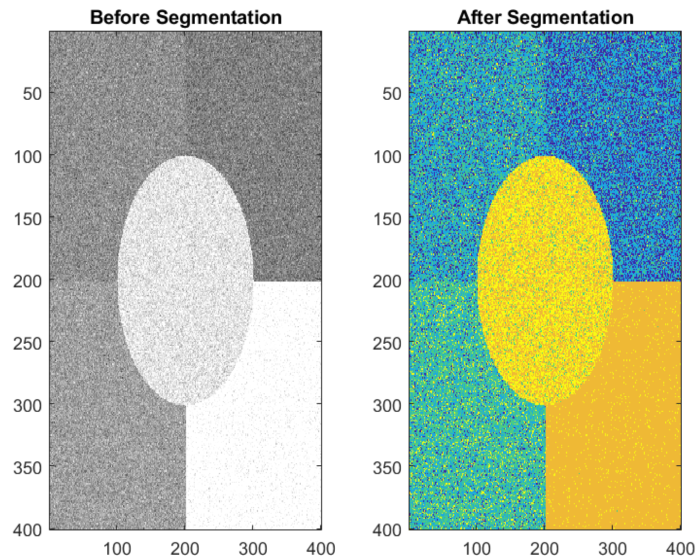

Figure 19: Figure 4.5: Synthetic image with 20% Gaussian noise before and after segmentation using k_means.

S= 33.8191%, SA= 47.2634%.

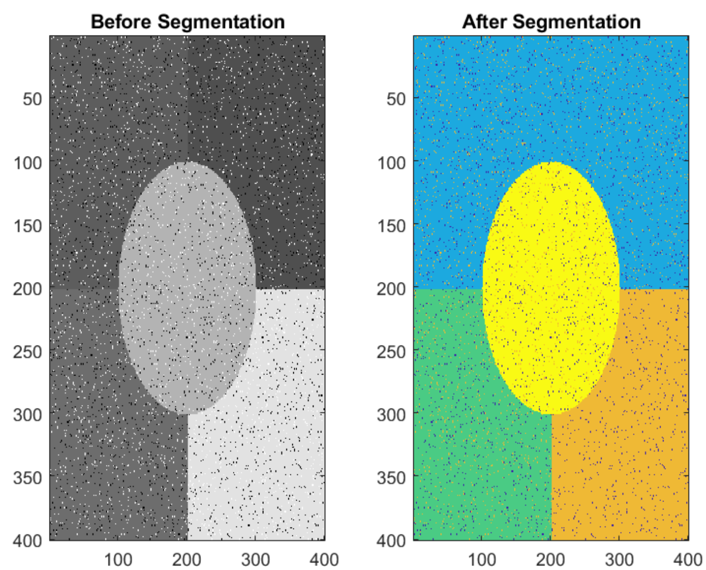

Figure 20: Synthetic image with 5% Salt and Pepper noise before and after segmentation using k_means.

S= 63.0018%, SA= 76.3422%.

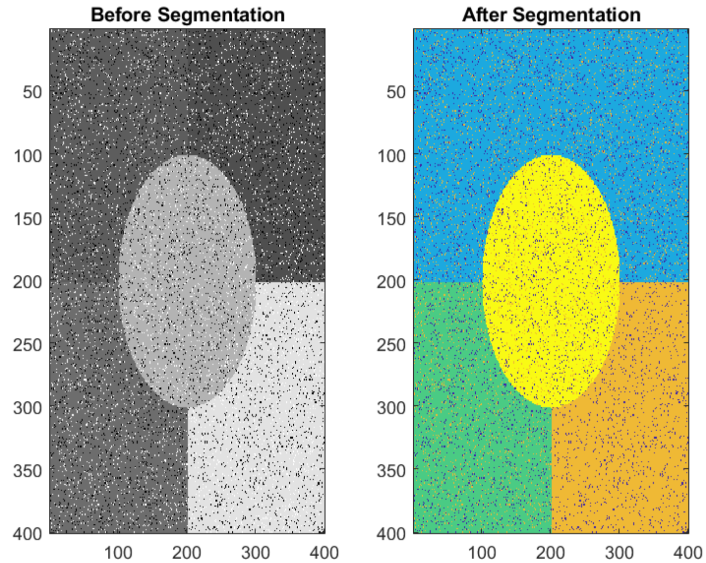

Figure 21: 4.7: Synthetic image with 10% Salt and Pepper noise before and after segmentation using k_means.

S= 59.5664%, SA= 72.7887%.

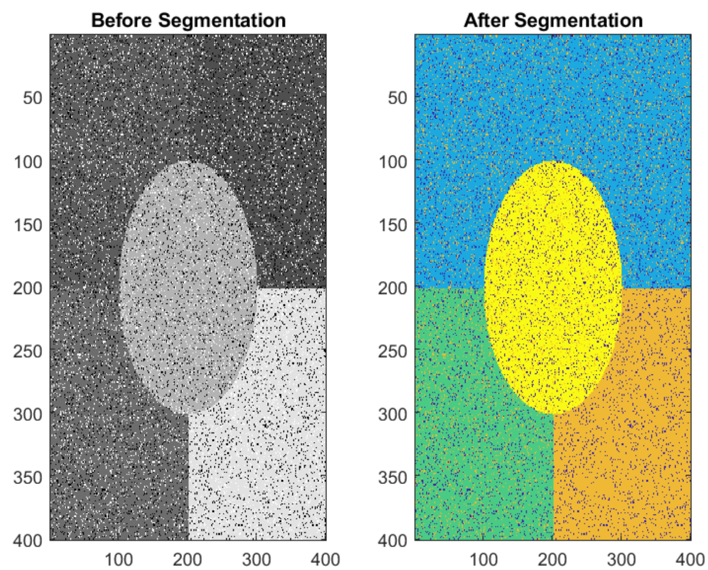

Figure 22: Synthetic image with 15% Salt and Pepper noise before and after segmentation using k_means.

S= 56.3224%, SA= 69.3696%.

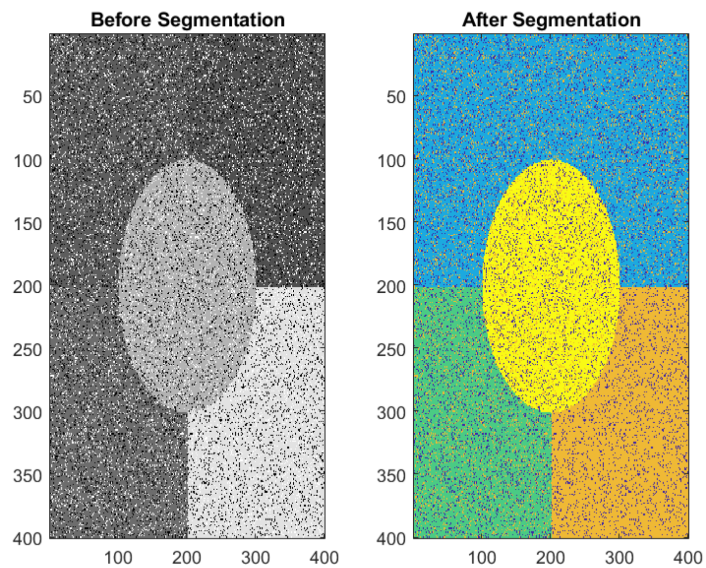

Figure 23: Synthetic image with 20% Salt and Pepper noise before and after segmentation using k_means.

S= 53.0049%, SA= 65.8248%.

The K-means algorithm performed better with the salt and pepper noise than the Gaussian noise, although the robustness of the algorithm wasn’t so good. The algorithm made 5 different clusters but couldn’t deal with the noise unfortunately.

2. Testing Fuzzy C-means Algorithm

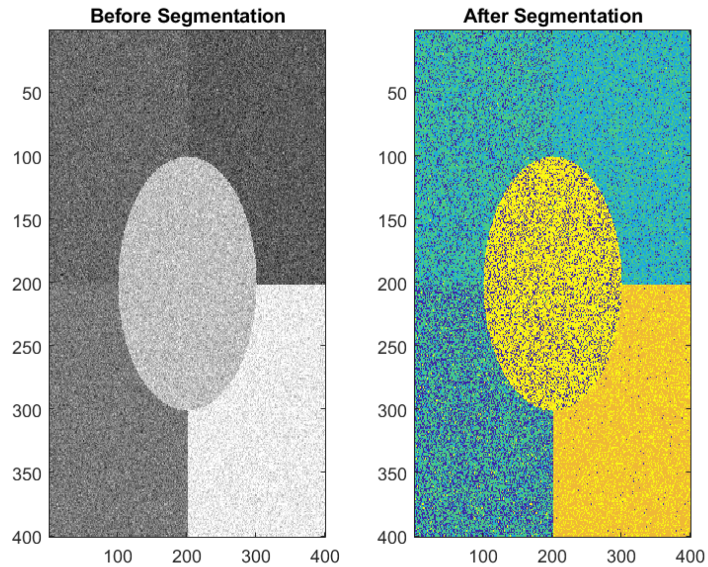

Figure 24: Synthetic image with 5% Gaussian noise before and after segmentation using Fuzzy C-means.

S= 34.9020%, SA= 47.3579%.

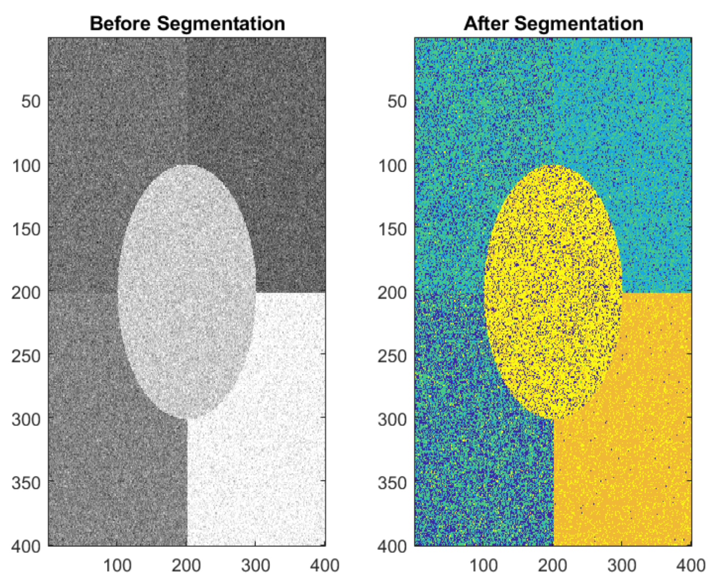

Figure 25: Synthetic image with 10% Gaussian noise before and after segmentation using Fuzzy C-means.

S= 35.6256%, SA= 47.8884%.

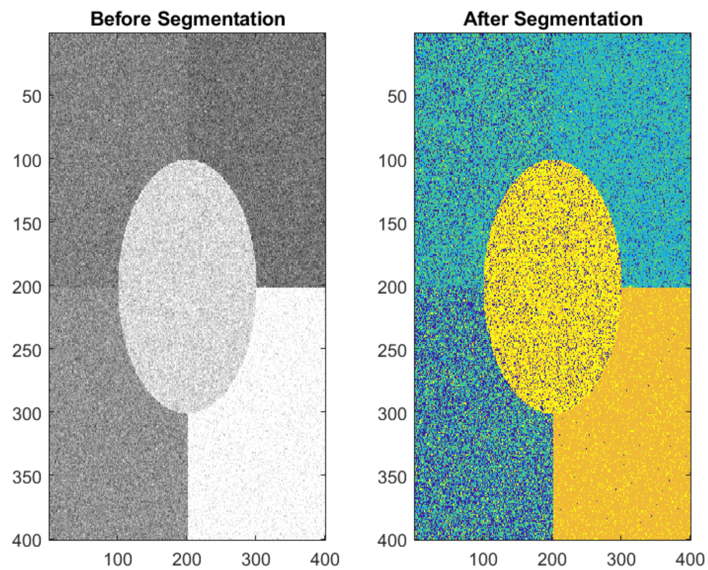

Figure 26: Synthetic image with 15% Gaussian noise before and after segmentation using Fuzzy C-means.

S= 35.2474%, SA= 47.6284%.

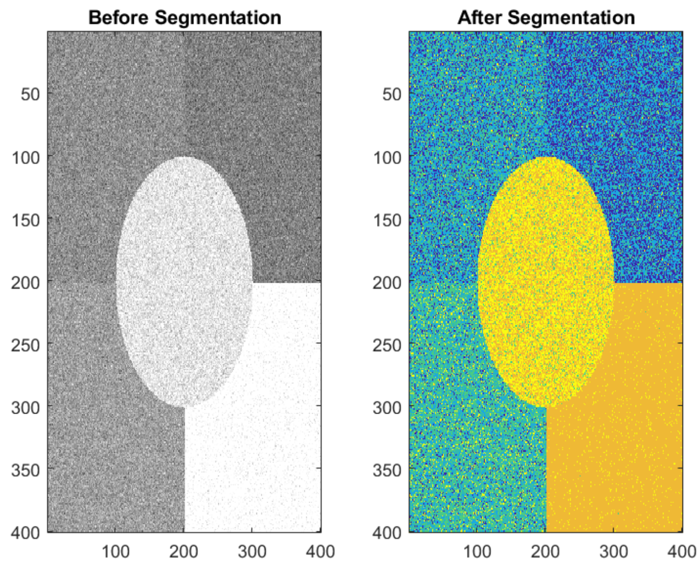

Figure 27: Synthetic image with 20% Gaussian noise before and after segmentation using Fuzzy C-means.

S= 33.8254%, SA= 46.9375%.

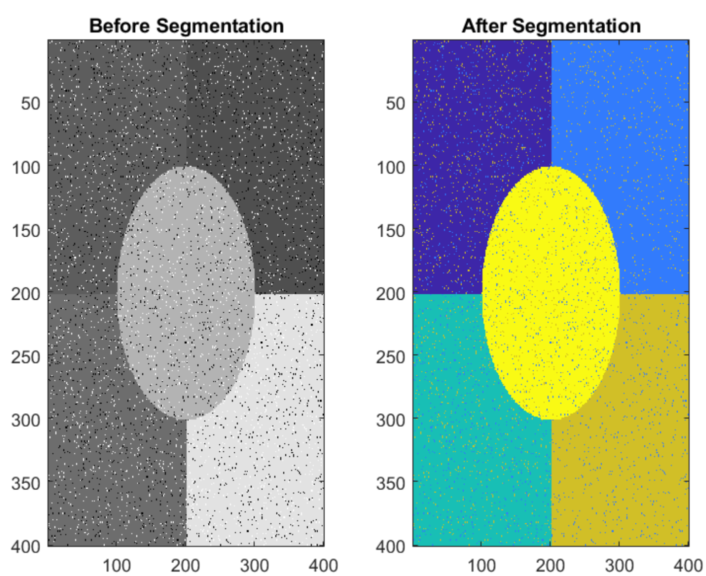

Figure 28: Synthetic image with 5% Salt and Pepper noise before and after segmentation using Fuzzy C-means.

S= 92.2767%, SA= 95.9832%.

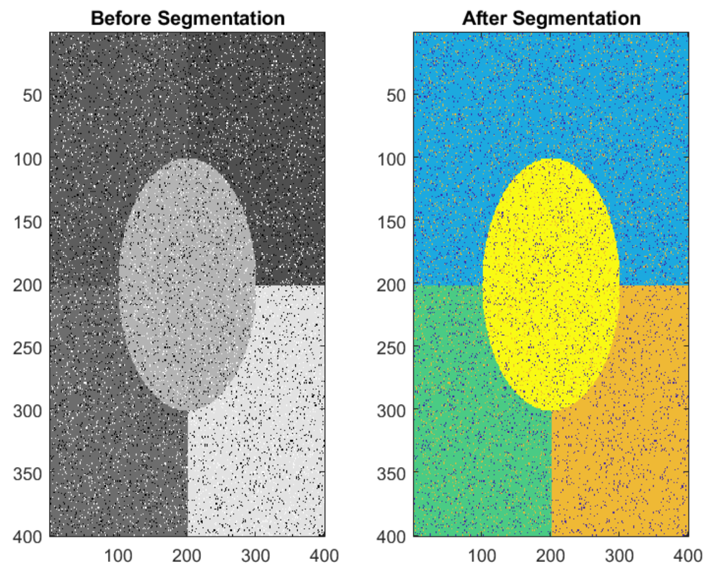

Figure 29: Synthetic image with 10% Salt and Pepper noise before and after segmentation using Fuzzy C-means.

S= 59.5664%, SA= 72.7887%.

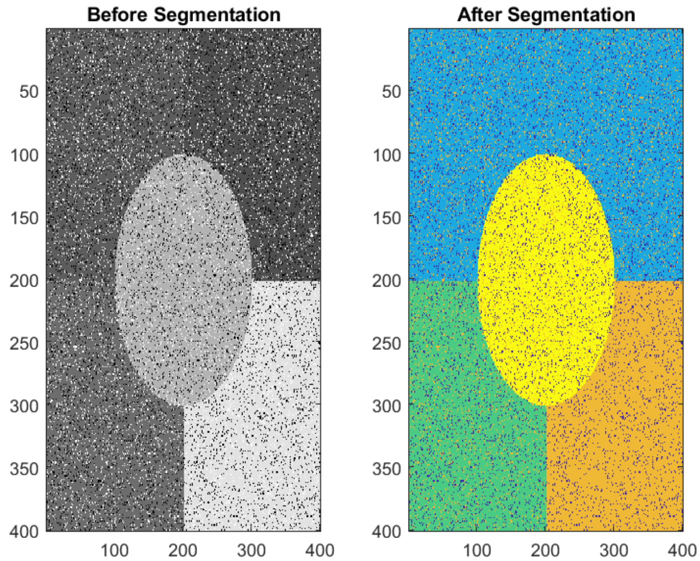

Figure 30: Synthetic image with 15% Salt and Pepper noise before and after segmentation using Fuzzy C-means.

S= 56.3224%, SA= 69.3696%.

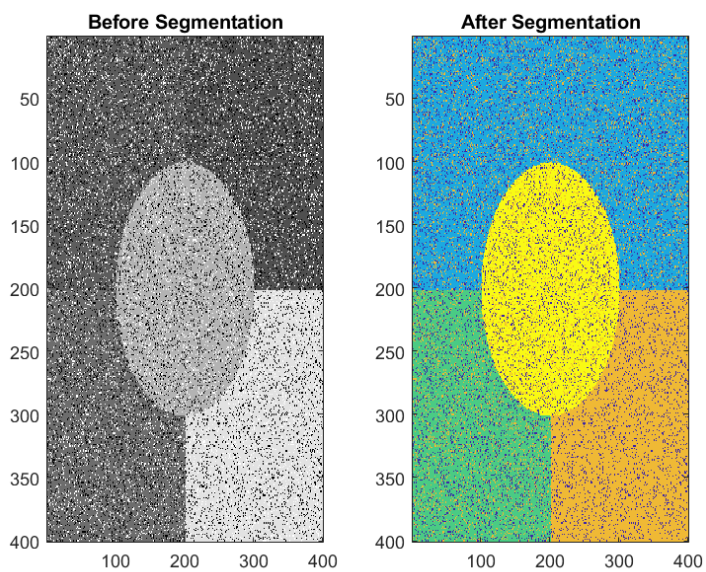

Figure 31: Synthetic image with 20% Salt and Pepper noise before and after segmentation using Fuzzy C-means.

S= 53.0049%, SA= 65.8248%.

Same as K-means, we got a better result from segmenting images with salt and pepper noise than with Gaussian noise. Segmenting the synthetic image with 5% salt and pepper noise resulted in S and SA equals 92.3 and 96% respectively, which are really good results, but continuing in raising the magnitude of the noise showed that the robustness isn't that great, and the algorithm needs more extra steps to perform well.

3. Testing Superpixel Fast Fuzzy C-means Algorithm

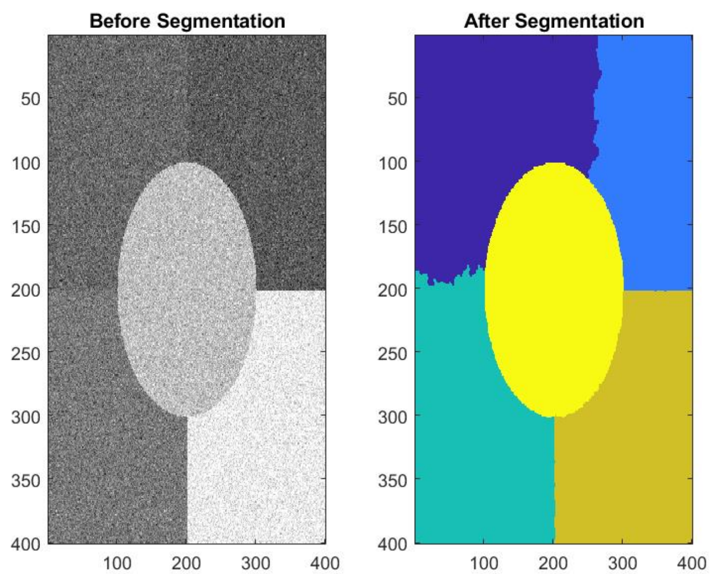

Figure 32: Synthetic image with 5% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 90.7349%, SA= 95.1424%.

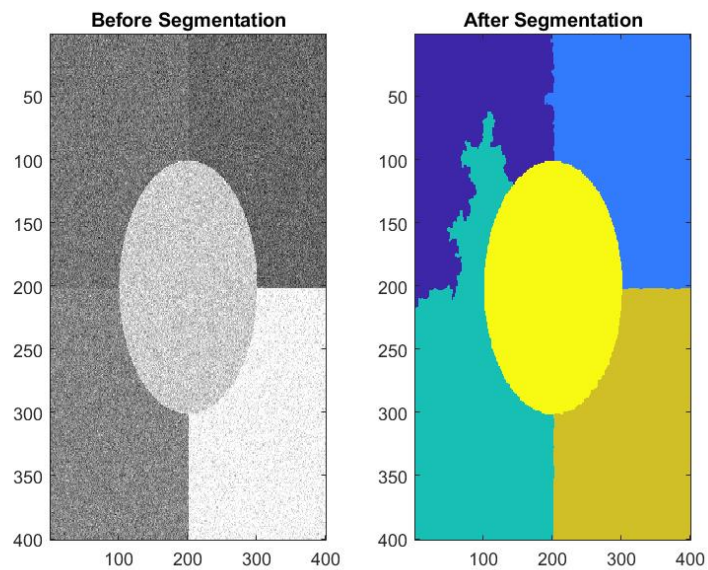

Figure 33: Synthetic image with 10% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 92.1331%, SA= 95.9055%.

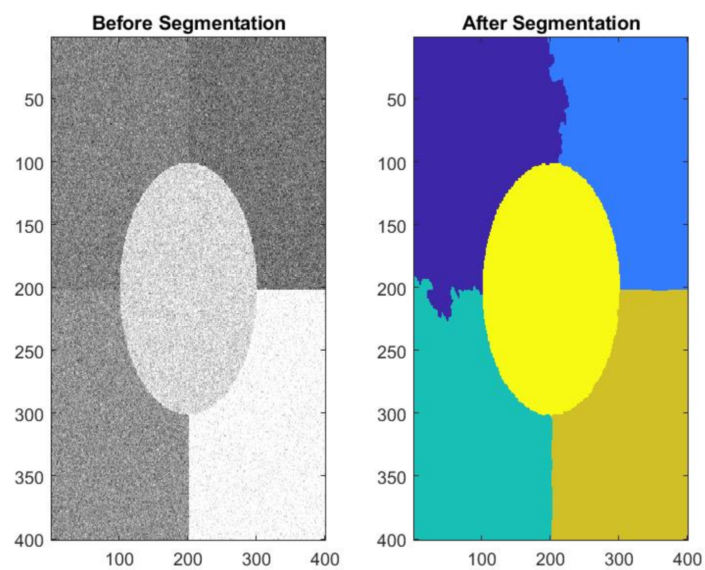

Figure 34: Synthetic image with 15% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 96.9189%, SA= 98.4353%.

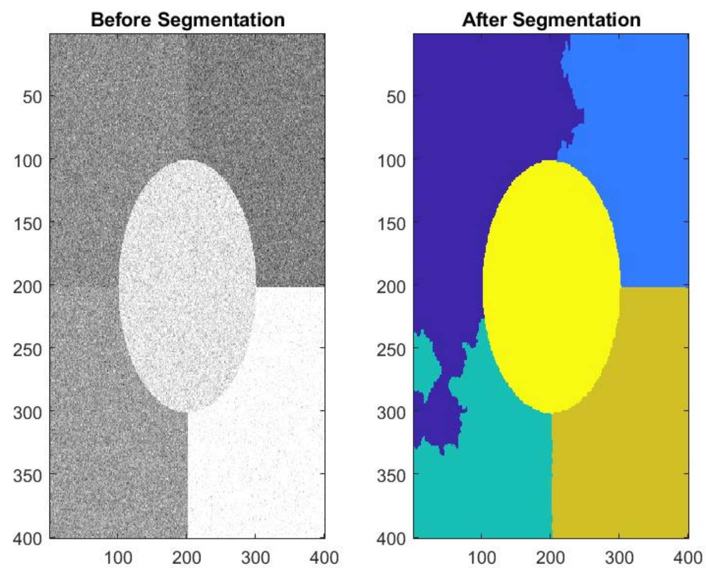

Figure 35: Synthetic image with 20% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 86.8063%, SA= 92.9372%.

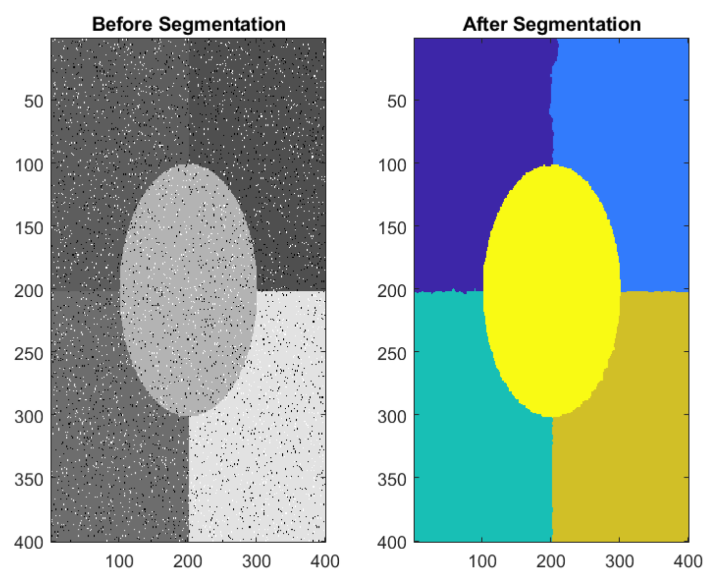

Figure 36: Synthetic image with 5% Salt and Pepper noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 99.0764%, SA= 99.5361%.

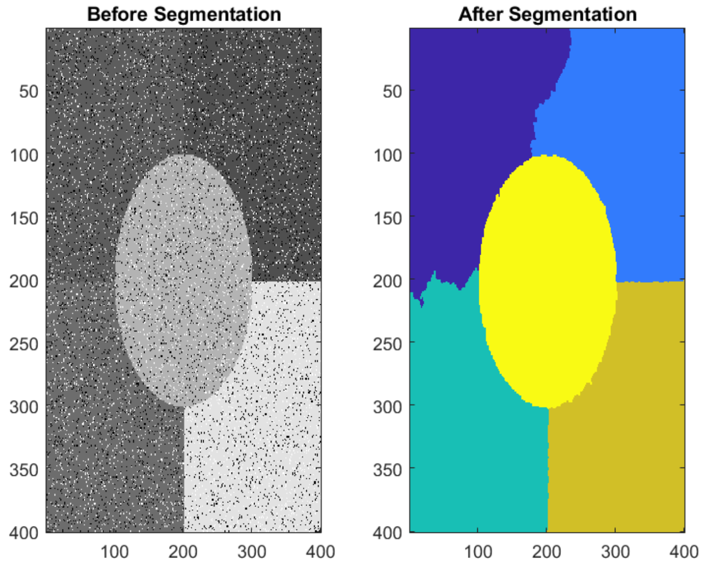

Figure 37: Synthetic image with 10% Salt and Pepper noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 96.1191%, SA= 98.0212%.

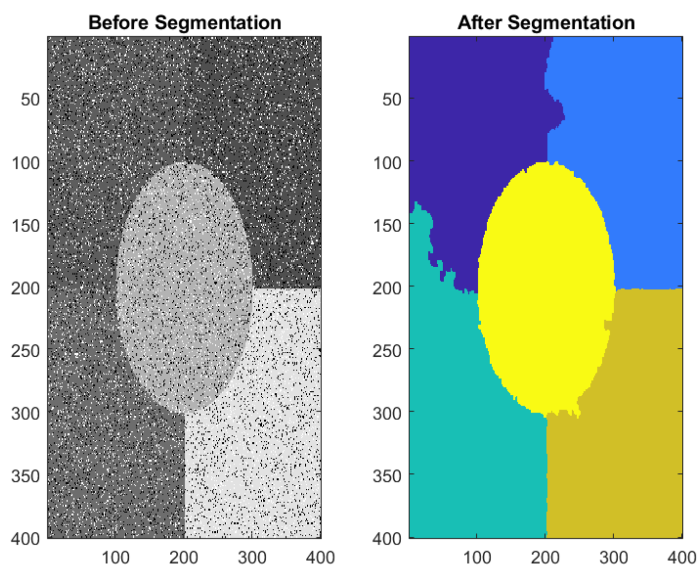

Figure 38: Synthetic image with 15% Salt and Pepper noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 94.9410%, SA= 97.4049%.

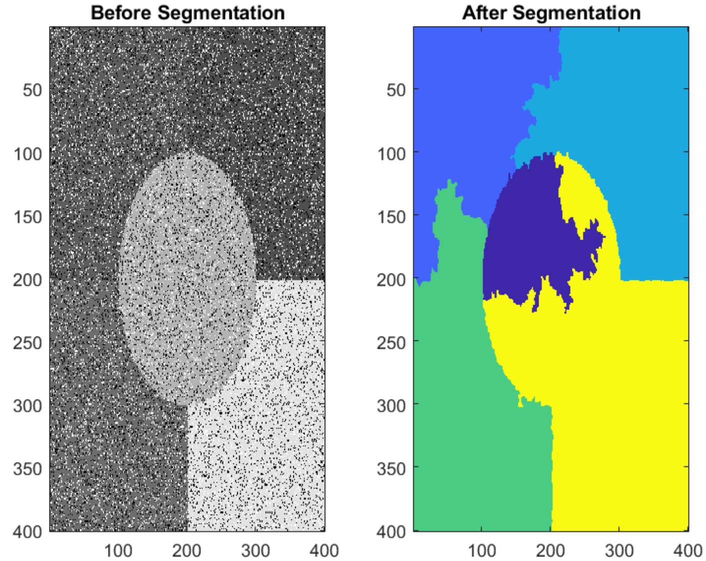

Figure 39: Synthetic image with 20% Salt and Pepper noise before and after segmentation using Superpixel Fast Fuzzy C-means.

S= 53.9346%, SA= 67.2477%.

As we can see from the S and SA results, Superpixel Fast Fuzzy C-means is superior to the previous methods (K-means and Fuzzy C-means). The result showed that the segmentation optimal accuracy is far exceeds the previous methods, and only showed poor performance when dealing with 20% salt and pepper noise.

D. The Data Set

The data contains three different varieties of chest cancer, including adenocarcinoma, large cell carcinoma, and squamous cell carcinoma, as well as one folder for normal cells. The photos are not in dcm format; instead, they are in jpg or png to suit the model. The primary folder that houses all the step folders is called the data folder. inside Data folder are test , train , valid.

1- Adenocarcinoma: The most prevalent kind of lung cancer, responsible for around 40% of all instances of non-small cell lung cancer and 30% of all cases overall, is lung adenocarcinoma. Breast, prostate, and colorectal malignancies all often become adenocarcinomas. The glands that release mucus and aid in breathing are located in the outer part of the lung where adenocarcinomas of the lung may be detected. Coughing, hoarseness, weakness, and weight loss are among symptoms.

2- Large-cell undifferentiated carcinoma is a kind of lung cancer that may develop anywhere in the lung and spread swiftly. 10 to 15 percent of all NSCLC patients often include this particular kind of lung cancer. Large-cell undifferentiated carcinoma has a propensity for rapid growth and dispersal.

3- Squamous cell carcinoma: This kind of lung cancer develops in one of the major branches of the airways or in the center of the lung, where the bigger bronchi connect the trachea to the lung. Approximately 30% of all non-small cell lung cancers are squamous cell lung cancers, and smoking is often a risk factor.

4- Normal: Those are normal CT lung scans.

E. Testing with real data, Lung CT Images

As we explained in the section before, we have four types of CT images, Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and Normal images. We will test the three algorithms using one normal image, and will add Gaussian noise in four different magnitudes, 5,10,15, and 20%. We used Gaussian because usually CT images get noisy with this kind of noise. Finally, we will compare the performance using the run time, along with eye inspection.

1. Testing K-means Algorithm

We did a lot of testing, and to get the best segmentation for the lungs using this algorithm, putting the clusters = 3 was the best option.

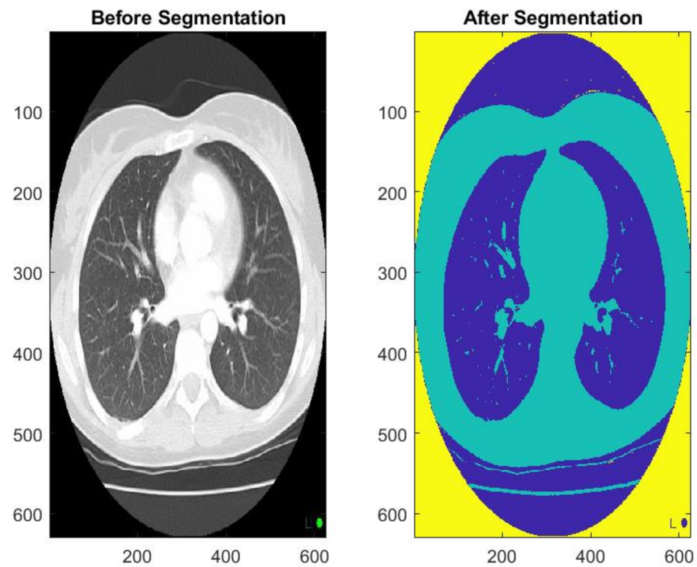

Figure 40: Lung CT image with 0% Gaussian noise before and after segmentation using K-means clustering algorithm.

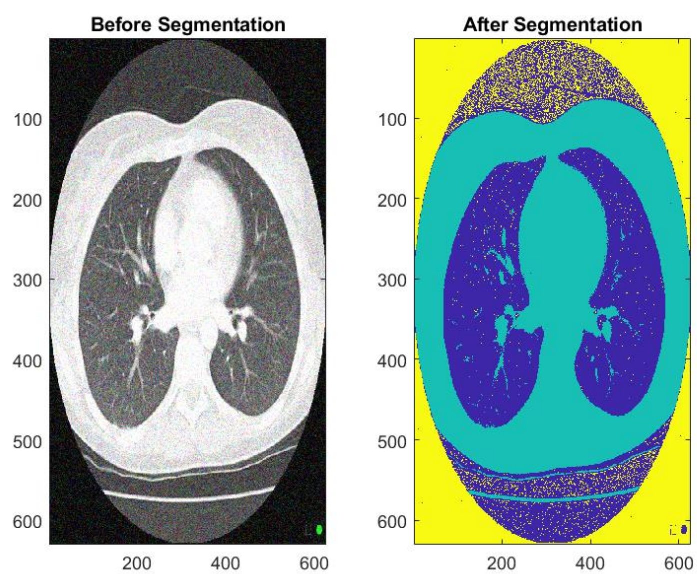

Figure 41: Lung CT image with 5% Gaussian noise before and after segmentation using K-means clustering algorithm.

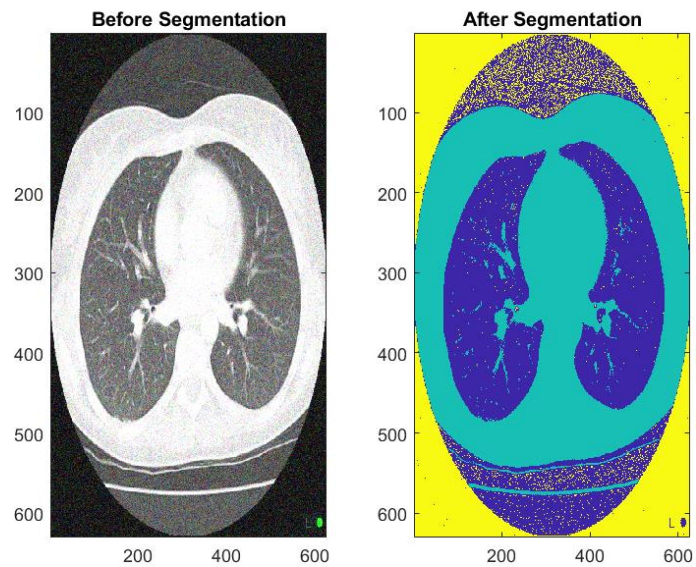

Figure 42: Lung CT image with 10% Gaussian noise before and after segmentation using K-means clustering algorithm.

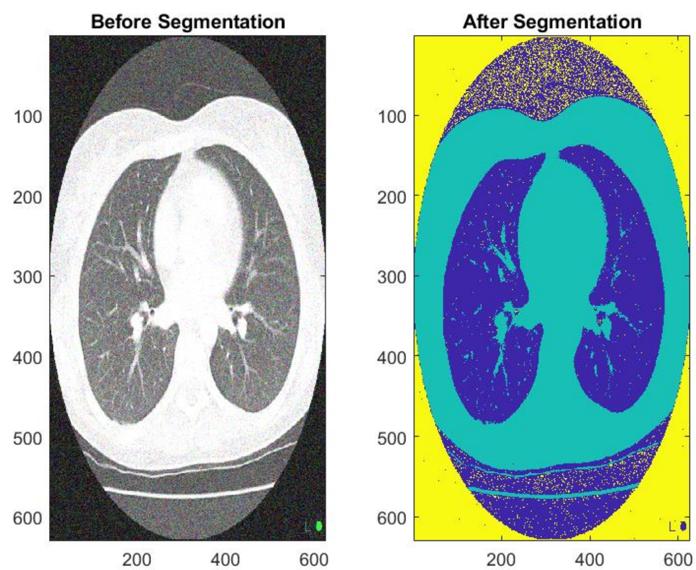

Figure 43: Lung CT image with 15% Gaussian noise before and after segmentation using K-means clustering algorithm.

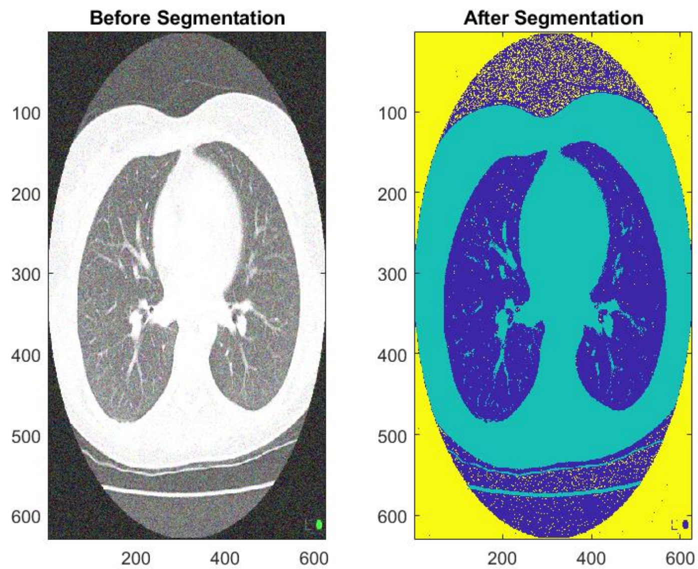

Figure 44: Lung CT image with 20% Gaussian noise before and after segmentation using K-means clustering algorithm.

The K-means clustering did a fine job with the first image with no noise at all, and segmented the lung in a good way, although after the noise addition, as we can see that the lung segment is noisy, so the algorithm can withstand a bit of adjustment to increase its robustness.

2. Testing Fuzzy C-means Algorithm

For this algorithm and the next one (Superpixel Fast Fuzzy C-means), we will use the clusters =4.

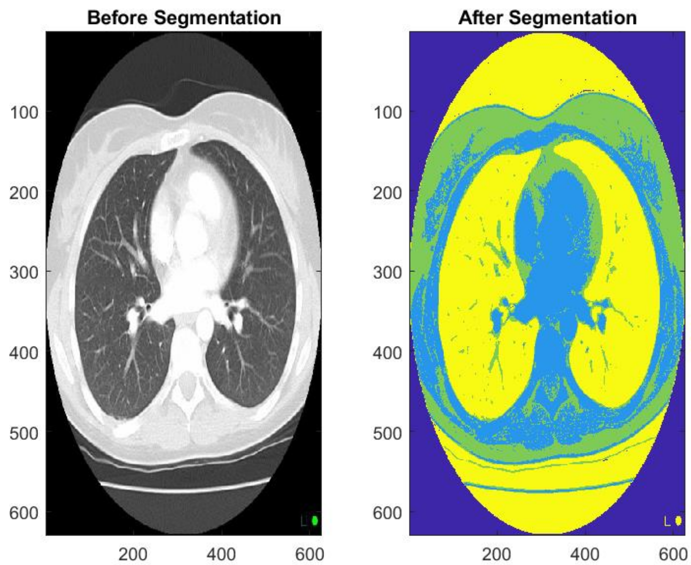

Figure 45: Lung CT image with 0% Gaussian noise before and after segmentation using Fuzzy C-means clustering algorithm.

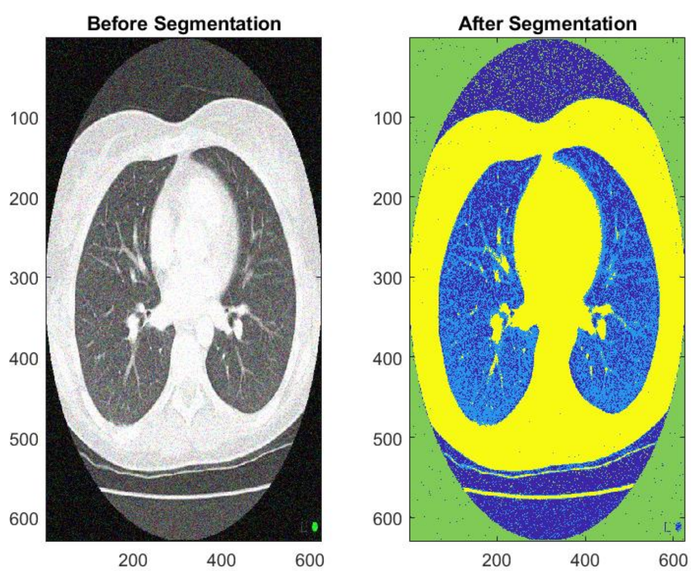

Figure 46: Lung CT image with 5% Gaussian noise before and after segmentation using Fuzzy C-means clustering algorithm.

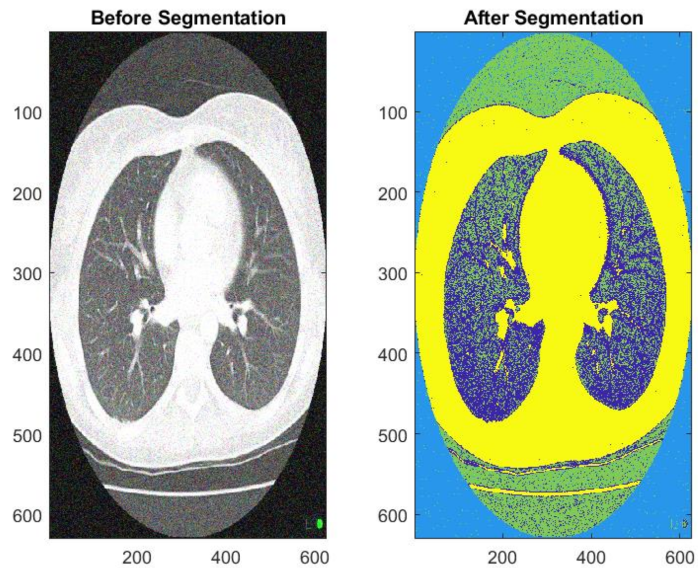

Figure 47: Lung CT image with 10% Gaussian noise before and after segmentation using Fuzzy C-means clustering algorithm.

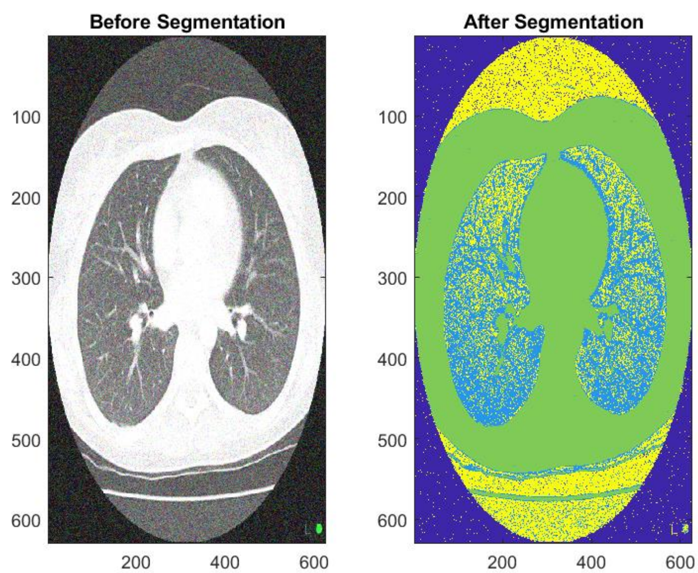

Figure 48: Lung CT image with 15% Gaussian noise before and after segmentation using Fuzzy C-means clustering algorithm.

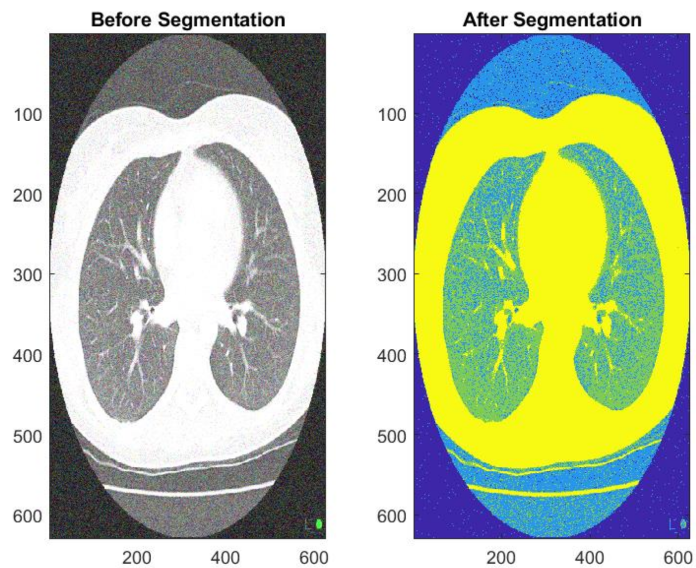

Figure 49: Lung CT image with 20% Gaussian noise before and after segmentation using Fuzzy C-means clustering algorithm.

As we saw in the previous algorithm (K-means clustering), this algorithm showed good lung segmentation without adding the noise to the image, but the more noise we added, it showed that this algorithm doesn't have good robustness, and it also needs some modification.

3. Testing Superpixel Fast Fuzzy C-means Algorithm

As we mentioned before, we will put the clusters = 4 in this algorithm.

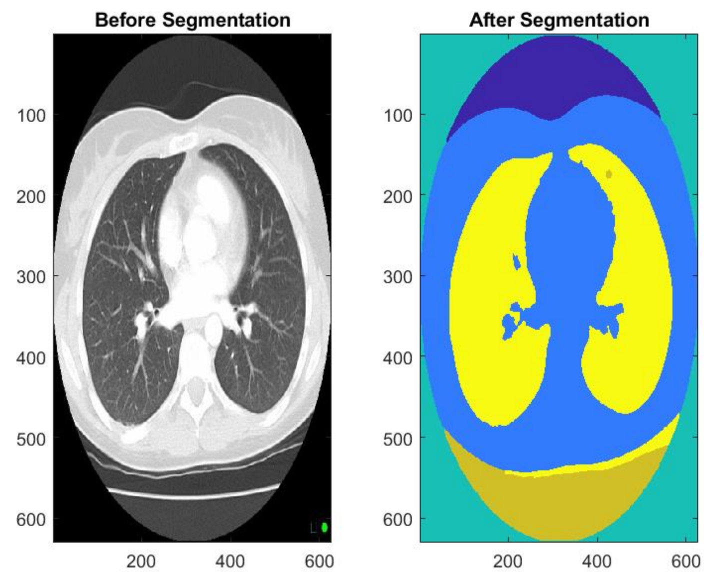

Figure 50: Lung CT image with 0% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means clustering algorithm.

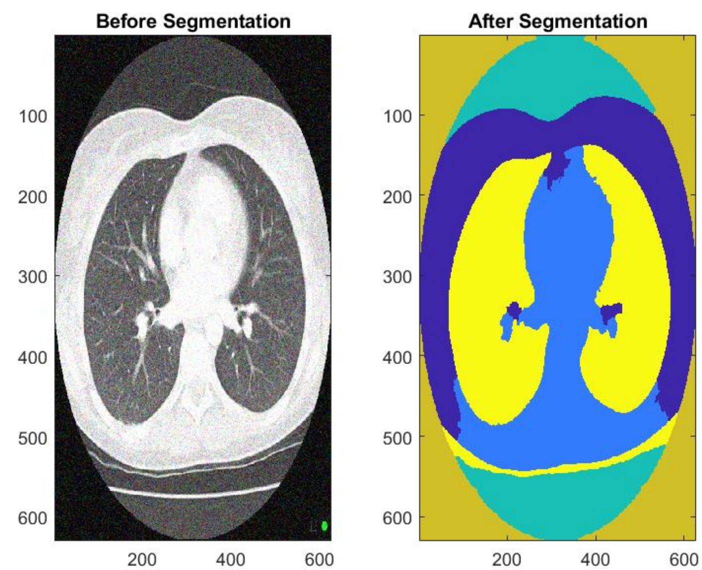

Figure 51: Lung CT image with 5% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means clustering algorithm.

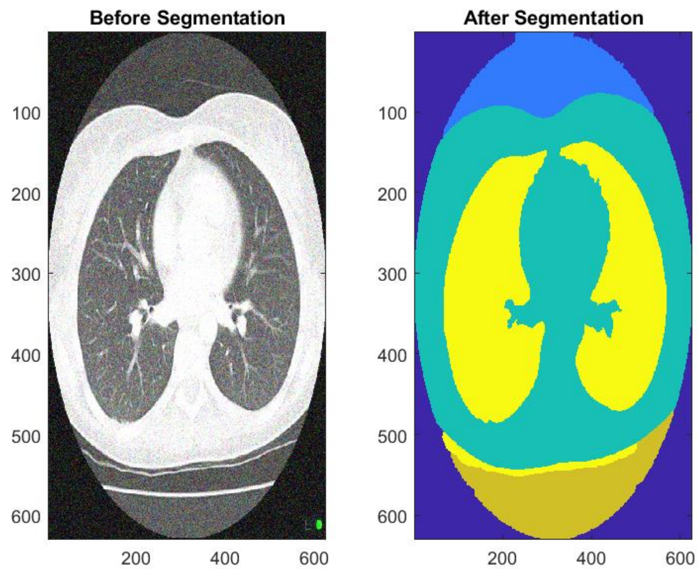

Figure 52: Lung CT image with 10% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means clustering algorithm.

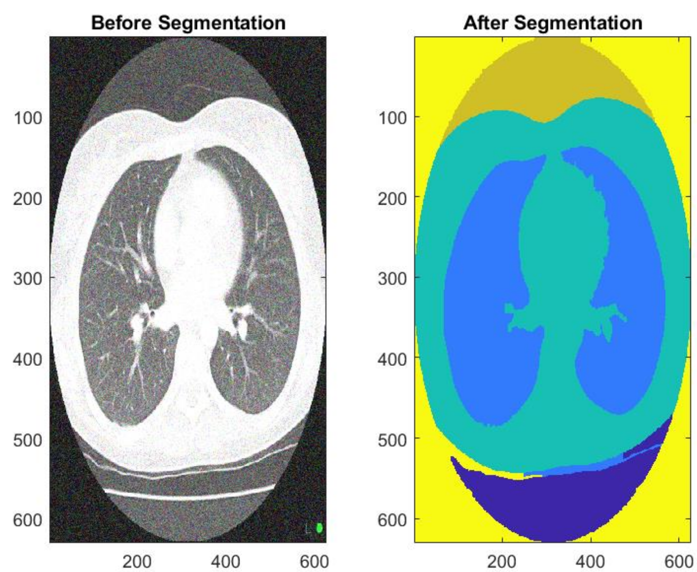

Figure 53: Lung CT image with 15% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means clustering algorithm.

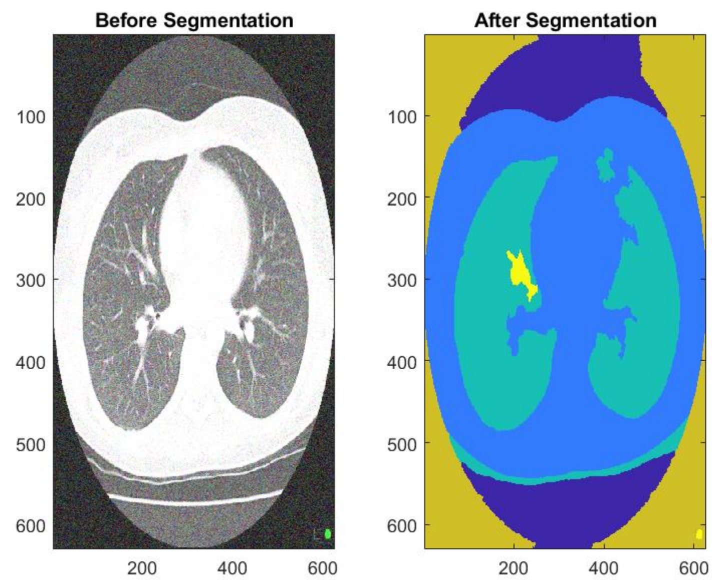

Figure 54: Lung CT image with 20% Gaussian noise before and after segmentation using Superpixel Fast Fuzzy C-means clustering algorithm.

As we can see from the comparison above, this algorithm has much more robustness than the previous two algorithms (K-means and Fuzzy C-means), it showed great segmentation results for the lung with the free noisy image and the noisy images, and we can clearly say that this is the best algorithm out of the three algorithms.

V. CONCLUSION AND FUTURE WORK

In chapter 4 (Results), we applied the three algorithms (Superpixel Fast Fuzzy C-means, Fuzzy C-means, and K-means) on two types of images:

1- The synthetic image (see figure 15) with 5 different grey levels.

2- The Lung CT image

We added two types of noise with different magnitudes to the synthetic image, which were salt and pepper, and Gaussian noise. However, we only added Gaussian noise to the lung CT image, because it’s the most common type of noise that affects CT images.

And from testing on the synthetic image, we were able to calculate the optimal segmentation accuracy (SA), and the degree to which the segmented picture and the ground truth are identical; the quantitative score (S), for the three algorithms mentioned above. Below, Table 4 shows the (S) scores, and Table 5 shows the (SA) scores.

Noise	K-means (%)	Fuzzy C-means (%)	Superpixel Fast Fuzzy C-means (%)
Gaussian 5%	35.6030	34.9020	90.7349
Gaussian 10%	35.5600	35.6256	92.1331
Gaussian 15%	34.6780	35.2474	96.9189
Gaussian 20%	33.8191	33.8254	86.8063
Salt & Pepper 5%	63.0018	92.2767	99.0764
Salt & Pepper 10%	59.5664	59.5664	96.1191
Salt & Pepper 15%	56.3224	56.3224	94.9410
Salt & Pepper 20%	53.0049	53.0049	53.9346

Table 4: The quantitative score (S)

Noise	K-means (%)	Fuzzy C-means (%)	Superpixel Fast Fuzzy C-means (%)
Gaussian 5%	48.2976	47.3579	95.1424
Gaussian 10%	47.8971	47.8884	95.9055
Gaussian 15%	47.6546	47.6284	98.4353
Gaussian 20%	47.2634	46.9375	92.9372
Salt & Pepper 5%	76.3422	95.9832	99.5361
Salt & Pepper 10%	72.7887	72.7887	98.0212
Salt & Pepper 15%	69.3696	69.3696	97.4049
Salt & Pepper 20%	65.8248	65.8248	67.2477

Table 5: The optimal segmentation accuracy (SA)

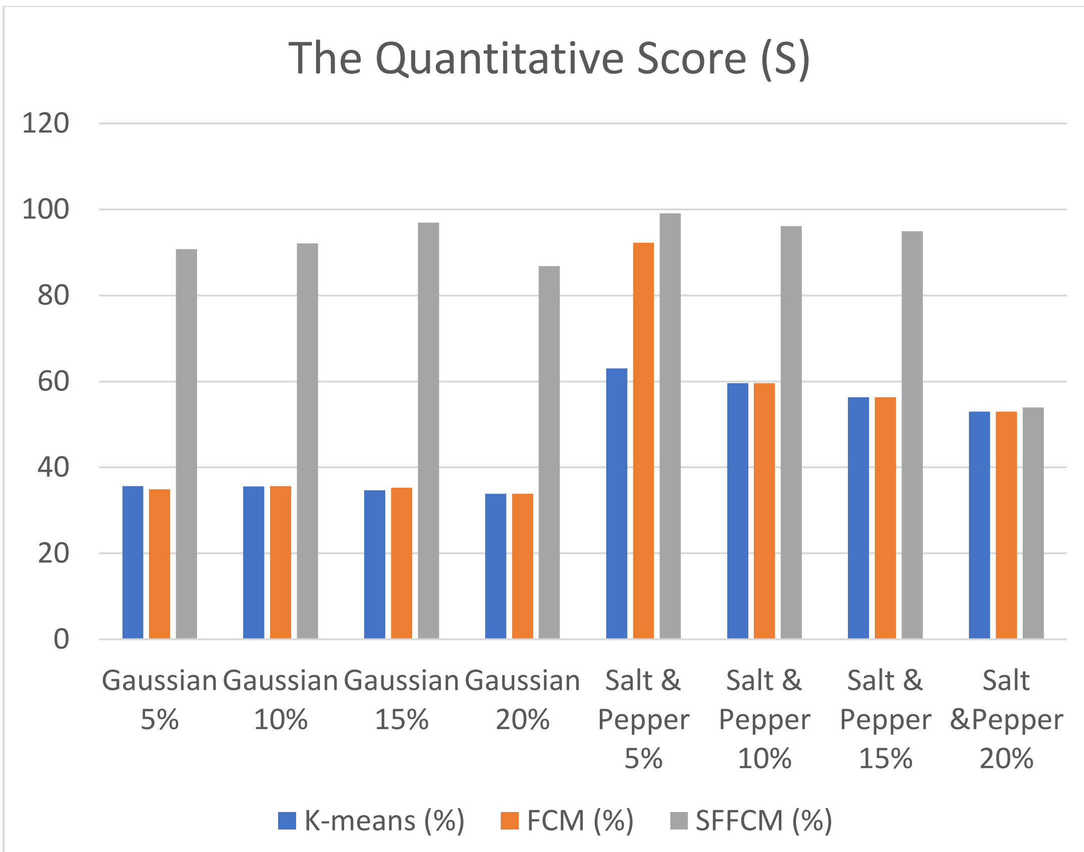

Figure 55: The quantitative score (S) for the three algorithms.

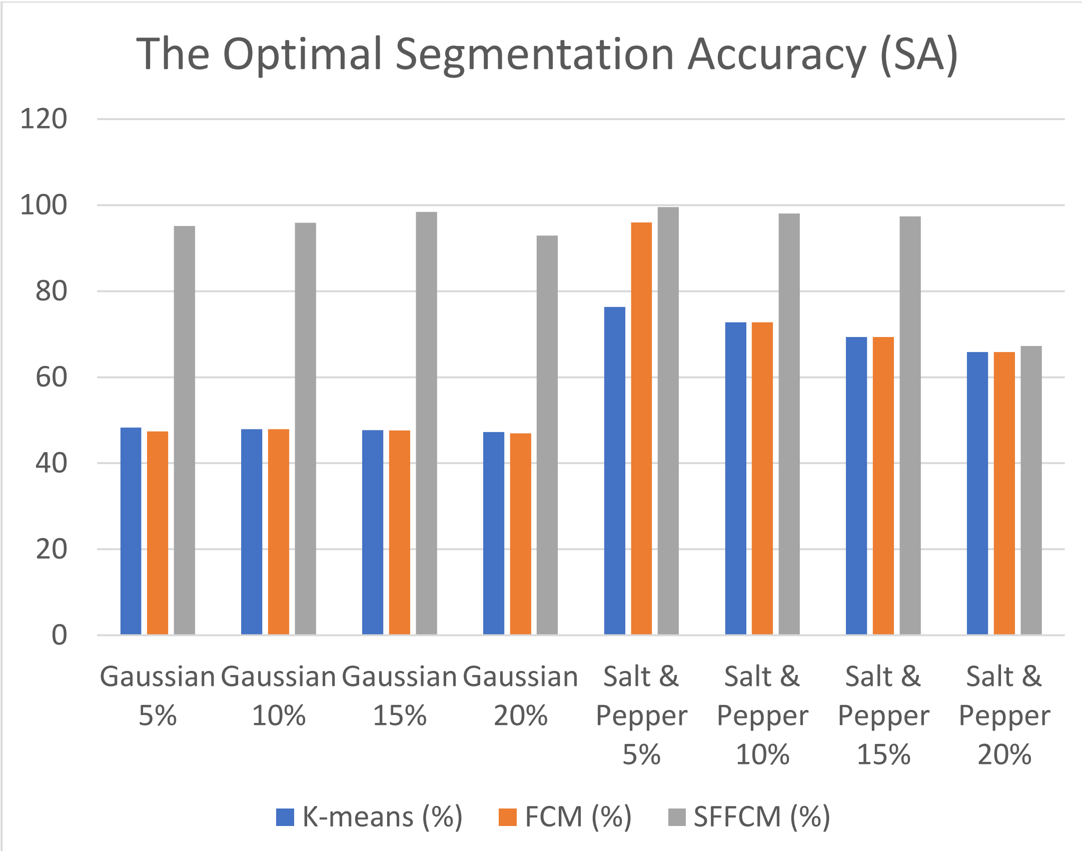

Figure 56 : The optimal Segmentation accuracy (SA) for the three algorithms.

The average (S) score when adding Gaussian noise for each algorithm is: 34.92% for K-means, 34.90% for Fuzzy C-means, and 91.65% for Superpixel Fast Fuzzy C-means. The average (S) score when adding Salt and pepper noise for each algorithm is: 57.97% for K-means, 65.29% for Fuzzy C-means, and 86.02% for Superpixel Fast Fuzzy C-means.

The average (SA) score when adding Gaussian noise for each algorithms is: 47.78% for K-means, 47.45% for Fuzzy C-means, and 95.61% for Superpixel Fast Fuzzy C-means. The average (SA) score when adding Salt and pepper noise for each algorithms is: 71.08% for K-means, 75.99% for Fuzzy C-means, and 90.55% for Superpixel Fast Fuzzy C-means.

Lastly, overall average of (S) for the three algorithms is: 46.44% for K-means, 50.01% for Fuzzy C-means, and 88.83% for Superpixel Fast Fuzzy C-means. Overall average of (SA) for all three algorithms is: 59.43% for K-means, 61.72% for Fuzzy C-means, and 93.08% for Superpixel Fast Fuzzy C-means. We conclude that K-means and Fuzzy C-means are almost the same when dealing with Gaussian noise, and Fuzzy C-means is better than K-means when dealing with salt and pepper noise, and Superpixel Fast Fuzzy C-means is far superior and has much more robustness in comparison with the previous two methods.

About the run time for the three algorithms when dealing with the lung CT image, the fastest is K-means in 0.38 seconds, the second is Superpixel-based Fuzzy C-means in 1.90 seconds, and the slowest is Fuzzy C-means in 3.37 seconds. Concerning the future work, we will develop an algorithm that will locate the nodules inside the lungs after the lung segmentation and classify the nodules into categories.

VI. REFERENCES

BOOKS

DOĞAN, Ö., & LALE, A. (2002). **A fuzzy algorithm for color quantization of images. Pattern Recognition.**

GARETH, J., DANIELA, W., TREVOR, H., & ROBERT, T. (2021). **An Introduction to Statistical Learning.**

JERROLD, T., J.ANTHONY, S., EDWIN, M. J., & JOHN, M. (2012). **The Essential Physics of Medical Imaging**. Philadelphia, PA 19103 USA: **LIPPINCOTT WILLIAMS & WILKINS, a WOLTERS KLUWER business.**

PAUL, S. (2009). **Fundamentals of Medical Imaging**. New York: **Cambridge University Press.**

RAFAEL, C. G., & RICHARD, E. W. (2008). **Digital Image Processing**. New Jersey: **Pearson Education, Inc.**

S.KEVIN, Z., HAYIT, G., & DINGGANG, S. (2017). **Deep Learning for Medical Image Analysis.**

VAGELIS, H. (2009). **Information Discovery on Electronic Health Records.**

ARTICLES

ALAM, M. J., SHAMS, N. A., & MD. ZUBAIR, H. (2020). A Robust CNN Framework with Dual Feedback Feature Accumulation for Detecting Pneumonia Opacity from Chest X-ray Images. **International Conference on Electrical and Computer Engineering (ICECE***)*.

ALBAN, N., LAURENT, B., ETIENNE, M., OUSMAN, B., MITHERAND, N., & MARTIN, N. (2017). Robust and fast segmentation based on fuzzy clustering combined with unsupervised histogram analysis. **IEEE Intelligent Systems**.

BHAT, S., R, S., KUMAR, S., & K, G. (2020). Convolutional Neural Network approach for the Classification and Recognition of Lung Nodules. **Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA-2020).**

CHONG, W., LE, Z., HOUWANG, Z., & HONG, Y. (n.d.). Improved superpixel-based fast fuzzy c-means clustering for image segmentation. **IEEE International Conference on Image Processing (*ICIP*).**

CHUNRAN, Y., YUANYUAN, W., & YI, G. (2018). Automatic Detection and Segmentation of Lung Nodule on CT Images. **International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018).**

GU, Y., LAI, Y., XIE, P., WEI, J., & LU, Y. (2019). MULTI-SCALE PREDICTION NETWORK FOR LUNG SEGMENTATION. **IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Venice, Italy.**

HEEWON, K., HYUNSOO, H., HYUNA, C., KIWON, S., & JIHYE, L. (2019). Pneumonia Detection with Weighted Voting Ensemble of CNN Models. **International Conference on Artificial Intelligence and Big Data.**

KITRUNGROTSAKUL, T., CHEN, Q., WU, H., IWAMOTO, Y., HU, H., ZHU, W., . . . CHEN, Y.-W. (2021). Attention-RefNet: Interactive Attention Refinement Network for Infected Area Segmentation of COVID-19. **IEEE Journal of Biomedical and Health Informatics.**

LUC, V. (1993). Morphological Gray scale Reconstruction in Image Analysis: Applications and Efficient Algorithms. **IEEE TRANSACTIONS ON IMAGE PROCESSING**, 2.

NAIK, A., EDLA, D. R., & KUPPILI, V. (2020). A combination of FractalNet and CNN for Lung Nodule Classification. **ICCCNT.**

NARSIMHA, R. K., & CHANDRAN, S. (2018). Segmentation on chest radiographs using Otsu's and K-means clustering methods. **Inventive Research in Computing Applications (ICIRCA 2018).**

RAJ, M. R., & SULOCHANA, C. H. (2014). An Efficient Lung Segmentation Approach for Interstitial Lung Disease. **International Conference on Circuit, Power and Computing Technologies [ICCPCT].**

SHAZIYA, H., SHYAMALA, K., & ZAHEER, R. (2018). Automatic Lung Segmentation on Thoracic CT Scans using U-Net Convolutional Network. **International Conference on Communication and Signal Processing, April 3-5, 2018, India. India.**

SILVANA, M., AKBAR, R., GRAVINA, H., & FIRDAUS. (2020). Optimization of characteristics using Artificial Neural Network for Classification of Type of Lung Cancer. **International Conference on Information Technology Systems and Innovation (ICITSI). Bandung.**

SUMAN, T., & AVI, M. (n.d.). Image Segmentation using k-means clustering,EM and Normalized Cuts. **University Of California - Irvine.**

TAO, L., XIAOHONG, J., YANNING, Z., SHIGANG, L., HONGYING, M., & ASOKE, K. (2019, September). Superpixel-Based Fast Fuzzy C-Means Clustering For Color Image Segmentationg. **IEEE TRANSACTIONS ON FUZZY SYSTEMS.**

VASILESCU, V., NEACŞU, A., CHOUZENOUX, E., PESQUET, J.-C., & BURILEANU, C. (2021). A DEEP LEARNING APPROACH FOR IMPROVED SEGMENTATION OF LESIONS RELATED TO COVID-19 CHEST CT SCANS. **IEEE 18th International Symposium on Biomedical Imaging (ISBI).**

VENKATACHALAM, D., DR PURANDHAR, R., AMUDHAN, M., RAGURAMAN, A., & MOHAN, E. (2021). An implementation of K-means clustering for efficient image segmentation. **10th IEEE International Conference on Communication Systems and Network Technologies.**

XU, R., WANG, Y., TIANTIAN LIU, X. Y., LIN, L., CHEN, Y.-W., KIDO, S., & TOMIYAMA, N. (2020). BG-Net: Boundary-Guided Network for Lung Segmentation on Clinical CT Images. **International Conference on Pattern Recognition (ICPR).**

YADAVENDRA, & CHAND, S. (2021). Pneumonia lung opacity detection and segmentation in chest x-rays by using transfer learning of the Mask R-CNN. **International Conference on Advanced Computing and Communication Systems (ICACCS).**

ELECTRONIC SOURCES

AL, J. E. (n.d.). Thresholding. Retrieved February 6, 2022, from https://www.cse.unr.edu/~bebis/CS791E/Notes/Thresholding.pdf

Pulmonary Nodules. (n.d.). (University of Rochester Medical Center) Retrieved January 2022, from https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=22&contentid=pulmonarynodules

OTHER SOURCES

IZA, S., NOOR, K., & SITI, S. (2020, January). Comparison of Different Image Segmentation Techniques on MRI Image. Researchgate.

RESUME

Ghaith Abdul Mohsen Tayara

Objective

Seeking a challenging position in the biomedical engineering field to enrich my knowledge, skills, and experience, and contribute to make a significant change in the field. I'm looking for an organization where I could help by using all my knowledge, skills, and experience.

Experience

Echo medical supplies (EMS) **February 11th 2018 – September 2019**

- Medical Sales Representative |
  - Build strong relationships with internal and external stakeholders and ensure meeting the expectations.
  - Coordinate meetings with doctors and hospital medical staff, which may include pre-arranged appointments or regular 'cold' calling.
  - Demo and present to doctors, practice staff and nurses in surgeries, hospital.
  - Organize conferences for doctors and other medical staff.
  - Collect and analyze data about customers to find new leads and opportunities.
  - Maintain positive working spirit with medical staff and support administrative staff during urgent situations.
  - Win new customers, as well as develop long-term relationships with existing ones.
  - Monitor competitor activity and competitors' products.
  - Stay informed about the activities of health services in a particular area.

  Brands: Ultrasound systems (Alpinion), Patient monitors (Bionics), Surgical Instruments (ASC & MAHE).

First National Medical Services-Siemens Health Care Partner **February 5th - April 5th, 2017**

Field Service Engineer (Trainee):

- Installation and verification for new devices.
- Perform the preventive maintenance.
- Provide technical customer support.

Medical Imaging Devices: The First National Medical Services Training Sessions from Feb 5th to Feb 14th, 2017 (Certified) with the following course contents:

-X Ray & Cathlab Basics	-Ultrasound Basics
-Cyclotron Basics	-MRI Basics
-Radiotherapy Basics	-Sales and Management Basics
-Computed Tomography Basics	

Husam Nimer Drugstore at Abbott Devices **April 5th 2017 –June 5th 2017**

Field Service Engineer (Trainee):

· I helped install, make the required maintenance, and do the preventive maintenance to the Immunoassay analyzers, Chemistry analyzers and CBC devices.

· Made the regular meetings with lab technicians.

Education

Master Of Engineering | 2019 - 2022 | Istanbul Aydin University (Turkey)

· Major: Electrical and Electronics Engineering ·
· Accumulative Average: 3.31/4
· Thesis topic: Segmentation of Lung Computed Tomography Images.

Bachelor Of Engineering | 2017 | Yarmouk University (Jordan)

· Major: Biomedical Engineering. Accumulative Average: 72.3% · Rating: good.

Volunteering

Mentor At IEEE EMBS| Yarmouk University Student Chapter

· November 24th, 2016 - 2019

Chairman and Founder at IEEE EMBS | Yarmouk University Student Chapter

· May 29th, 2015 – November 23rd 2016

IEEE Jordan Conference on Applied Electric Engineering And Computing Technologies (AEECT)

· November 3rd -5th, 2015

Technical Skills

· MATLAB (Professional)
· Python (Beginner)
· Arduino (Programming and Design).
· Circuit maker and Proteus
· C++ and C# Programming
· Microsoft Office

Personal Skills

· Time Management
· Work under pressure
· leadership
· Problem solving
· Teamwork
· Sense of responsibility

Languages

· Arabic: Native.
· English: Fluent.