A Combination of K-Means and Fuzzy C-Means for Brain Tumor Identification

Cristy Atika Sari[1]*, Wellia Shinta[2], Hidayah Rahmalan[3]

[1]Department of Informatics Engineering, Faculty of Computer Sciences, Dian Nuswantoro University, Indonesia
[2]Department of Information System, Faculty of Computer Sciences, Dian Nuswantoro University, Indonesia
[3]Universiti Teknikal Malaysia Melaka, Malaysia

Abstract

Purpose: Magnetic Resonance Imaging is one of the health technologies used to scan the human body in order to get an image of an orgasm in the body. MRI imagery has a lot of noise that blends with the tumor object, so the tumor is quite difficult to detect automatically. In addition, it will be difficult to distinguish tumors from brain texture. Various methods have been carried out in previous studies.

Methods: This study combines the K-Means method and Fuzzy C-Means (FCM) to detect tumors on MRI. The purpose of the combination is to get the advantages of each algorithm and minimize weaknesses. The method used is Contrast Adjustment using Fast Local Laplacian, K-Means FCM, Canny edge detection, Median Filter, and Morphological Area Selection. The dataset is taken from www.radiopedia.org. Data taken were 73 MRI of the brain, of which 57 MRIs with brain tumors and 16 MRIs of normal brain Evaluation of research results will be calculated using Confusion Matrix.

Result: The accuracy obtained is 91.78%.

Novelty: K-Means method and Fuzzy C-Means (FCM) to detect tumors on MRI.

Keywords: MRI, Tumor Identification, Image Segmentation, K-Means, Fuzzy C-Means

Received March 2021 / **Revised** March 2021 / **Accepted** May 2021

INTRODUCTION

MRI or Magnetic Resonance Imaging is one of the technologies in the health sector that is used to scan the human body in order to obtain an image of an organ in the body. This technology uses a magnetic field and radio waves that act like sensors to get the structure of internal organs. MRI is often used to detect abnormalities in internal organs, one of which is the human brain. The purpose of this research was to detect tumors on MRI of the brain automatically, thus shortening the time to identify them and medical execution could be carried out immediately. However, MRI has a lot of noise, making it difficult to distinguish the tumor from the texture of the brain.

Various studies to detect brain tumors have been done before and the method that is often used is segmentation. Segmentation is the process of partitioning an image into several regions or classes based on its characteristic equation[1]. In previous paper by [2] Baid [3], Benson [4], and Padlina [5] it can be concluded that the use of K-means and FCM (Fuzzy C-Means) clustering methods is very good because in terms of program travel time is faster than those using the SVM, PSO or GA method and the results of image segmentation are accurate in detecting brain tumors with accuracy. an average of more than 88%, but must be accompanied by a noise removal step to reduce noise in the image. In addition, K-Means and FCM have other advantages in sensitivity. K-Means are more susceptible to local optima and outliers, which means they are more sensitive to color differences.

The advantage of FCM is that it is better at clustering convex shapes based on edges [4]. K-means performs the segmentation process faster than FCM. In previous research by Ghosh [6] and Cebeci [7], the two

studies compared the performance of K-Means and FCM. The result is that K-Means is better used on large data because the process is fast, but the results often change. Meanwhile, FCM is better used for data with a lot of noise, but the process is very long because it uses Fuzzy iterations.

From this problem, a research was conducted to combine the K-Means and FCM to minimize the disadvantages of each. Previously, MRI would be subjected to Fast Local Laplacian Filtering to clarify texture and contrast. Then the results are classified using the K-Means FCM. The classification results will be subjected to Canny edge detection to clarify the edges of the tumor. After that, the resulting lines are trimmed using a median filter and then selected using a morphological area selection to eliminate small objects that are considered noise or not tumors. After that the results of the detection will be marked with regionprops.

METHODS
The following is a rule of the proposed method as shown in Figure 1.
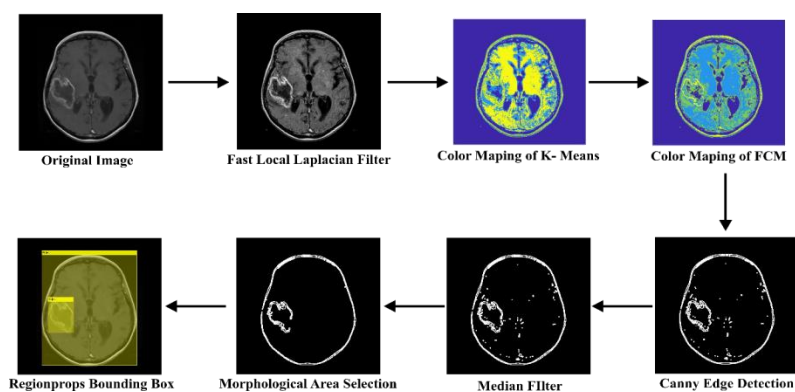

Figure 1. Proposed Method

Figure 1 shows the rule of the proposed method. The following is a brief explanation:
Step 1 : Imange Enhancement using Fast Local Laplacian to improve contrast and texture.
Step 2 : Then the image is converted to graycale.
Step 3 : Cluster image using K-Means and FCM. Color mapping is used so that you can see the difference.
Step 4 : Detect canny edges to define the edges on the image.
Step 5 : Median filter is used to smooth out the lines from Canny.
Step 6 : Morphological Area Selection is used to eliminate objects.
Step 7 : Regionprops is used as tumor marking.
For more details, the above steps will be described in more detail as follows.

Image Acquisition
MRI of the human brain was taken from the web www.radiopedia.org. with a total of 73 MRI images of which 16 images were normal brains, and 57 images were brain tumors diagnosed. The image taken is Axial T1 C + type and T2 type with RGB color with JPG / JPEG / PNG format. Then all the images will be converted into JPG format.

Image Enhancement
This stage is used to increase or decrease the image quality for further processing. At this stage, use a contrast adjustment to increase contrast and improve the texture of the image. Contrast is a very clear difference from things being compared, so that contras adjustment is a method of adjusting contrast to increase or decrease the quality of an image In this research, the algorithm used is Fast Local Laplacian filtering. Local laplacian uses two values, namely sigma and alpha, where the sigma value is used to process the detail of the image, and the alpha value is used to increase contrast. The sigma value cannot be negative and ranges from 0 (zero) to 1 (one). Sigma is used as a limit or measurement tool for alpha. In addition there are also rules for alpha, namely:
a. An alpha value of less than 1 will effectively increase the contrast so that the image becomes more detailed. In this condition, if the sigma value is greater than alpha, the result will be more contrast (detail). If the alpha value is closer to or greater than sigma, the result will be closer to the original image.

b. If alpha is equal to 1, the input and output images are the same.
c. If the alpha is more than 1, the contrast will be reduced so that the image becomes smoother. In this condition, the closer the sigma value is to zero, the result will look like the original image. If the sigma value approaches 1, the smoother the result will be.

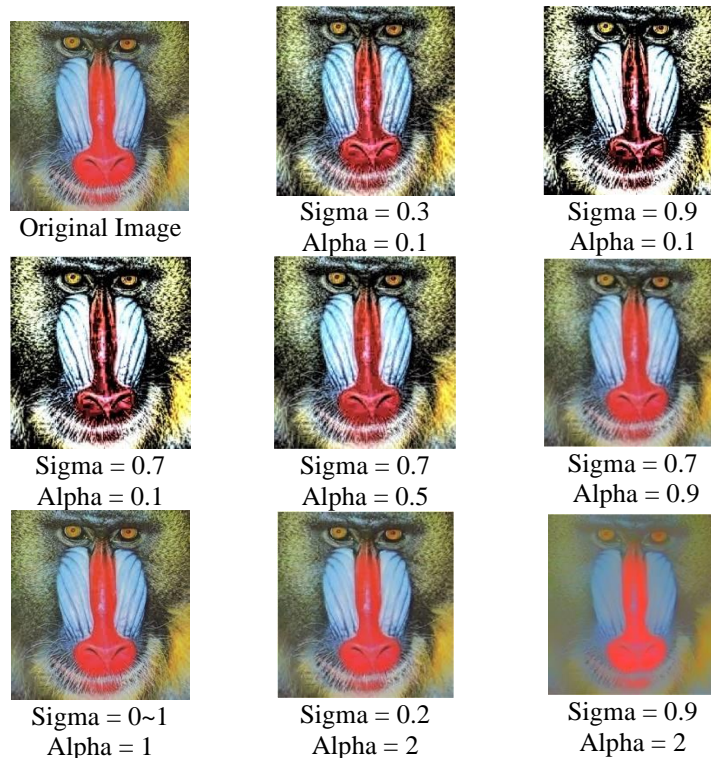For comparison, it can be seen from the image as shown in Figure 2.

Original Image	Sigma = 0.3 Alpha = 0.1	Sigma = 0.9 Alpha = 0.1
Sigma = 0.7 Alpha = 0.1	Sigma = 0.7 Alpha = 0.5	Sigma = 0.7 Alpha = 0.9
Sigma = 0~1 Alpha = 1	Sigma = 0.2 Alpha = 2	Sigma = 0.9 Alpha = 2

Figure 2. Fast local laplacian filter

In this study, the values of sigma = 0.5 and alpha = 0.4 were taken. This value is obtained from experimental results in the sigma range 0 to 1, and alpha 0 to 2

Convert Image to Grayscale
The first stage is changing the RGB image to Grayscale before moving on to the next process as shown in Figure 3.
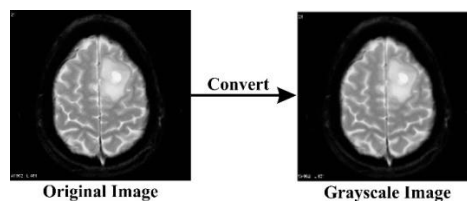
Figure 3. Convert to Grayscale

K-Means Clustering
This algorithm was first introduced by MacQueen and further developed by Hartigan and Wong, until it becomes a simple algorithm for data clustering [8] . K-Means is basically a partition method used to analyze and classify data by determining the location of the initial cluster center and the average distance between various data input points. The cluster center will continue to change until it gets the closest average distance [7]. In brief, the K-means clustering process has been describe [8] ;
1. Determine the random starting point of the k centroid cluster.
2. Calculate the distance of each pixel to the centroid, and classify it.
3. Calculate the new centroid value based on the member of each centroid.
4. Repeating steps 2 and 3, until the new centroid value does not change from the previous centroid value.

In this study using cluster value 4 (k = 4). This value is obtained from testing cluster values in the range 2 to 10, and the most results are at cluster value 4.

Fuzzy C-Means Clustering
The cluster results from K-Means are then processed by FCM. The performance of FCM is faster because the image has been clustered before by K-Means, while FCM only works for clusters that are not yet in accordance with its class. In this study, the number of FCM clusters used was 4 (k = 4). Fuzzy C-means is the result of minimizing the objective function of C-Means, the objective function is as follows:

$$J_m = \sum_{i=1}^{N} \sum_{j=i}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \tag{1}$$

For each x is a set of data points, c is the cluster center point, and m is the larger real value 1, $\|x\_i\text{-}c\_j\|^2$, is the distance between the data point and the cluster center point calculated by the Euclidean formula, Uij is the degree of membership of each data. The membership function of the FCM can be written as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} \tag{2}$$

The calculation stages of the FCM are as follows [5]:
1. Determine the cluster points, c1, c2, c3 and c4.
2. Calculating the value of the degree of membership of each data tested.
3. Renew the cluster center by applying the new formula as follows:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{3}$$

4. Update the value of the degree of membership of each data to the value of the new cluster center point. Repeating steps 2 and 3, if you find that the new cluster center value is different from the previous cluster.

Canny Edge Detection
After the clustering process, the results are processed with Canny edge detection which serves to detect and thicken the edges of the tumor. Following are the stages of Canny[9] [10]:
1. Reduce image noise using a gaussian filter.
2. Determine the gradient intensity and image direction. The gradient of each pixel can be determined using other edge detection operators such as Robert, Prewit, and Sobel. Meanwhile, determining the direction by mapping the angle into 4 parts, namely 0 °, 45 °, 90 °, and 135 °
3. Apply non-maximum supperession to eliminate the value of non-maximum pixels. The goal is to make the blur / smooth edges sharper.
Classify each pixel into 2 (two) classes, namely edge pixels and not edge pixels. The classification uses a hysteresis threshold with two thresholds, namely the upper threshold and the lower threshold.

Median Filter
Median filter is used to smooth Canny edge detection results. Median filter was chosen because it is more efficient in removing noise without affecting the surrounding pixels [8]. To calculate the median use the formula (4):

$$Median\ (m_x) = \begin{cases} x_{i(n+1)/2} & ; n\ odd \\ \frac{1}{2}\left[x_{i\left(\frac{n}{2}\right)} + x_{i\left(\frac{n}{2}\right)+1} \right] & ; n\ even \end{cases} \tag{4}$$

Where x is the value of each pixel that has been sorted and n is the pixel frequency. The following is the result of the median filter process after the image is enlarged or zoomed in as shown in Figure 4.
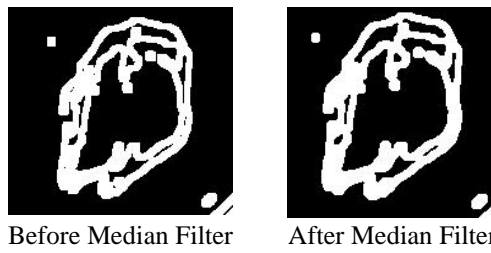
Before Median Filter   After Median Filter

Figure 4. Comparison of Before and After Median Filter

How to Reconciliaiton and Citations

This stage serves to eliminate small objects and fine lines that are considered noise or not tumors. Elimination is done based on the minimum object size that will not be eliminated. For the size of this morphological selection process, using the reference that the average size that can be of a brain tumor is 78.6mm (1D) or 4240.2mm2 (2D) or 34416.5mm3 (3D) [11].

To obtain the appropriate minimum size, an experiment was carried out based on the average tumor size so that elimination only leaves the tumor object. The best and safest number obtained in the image size ratio is 1: 350, or it can be written as 1/350.

RESULT AND DISCUSSION

The image dataset was taken from the internet, as many as 73 MRI of the brain, of which 57 MRI with brain tumors and 16 MRI of normal brain where the data is public data. The types of MRI used in this study are Axial weighting T1 C + and Axial T2. The image is formatted between JPG and PNG with RGB color. Each captured image will be converted into JPG format using an online converter from the internet without changing the pixel size of the image. The following are the types of MRI available on the radiopedia web as shown in Figure 5.

| Axial ADC | Axial DWI | Axial FLAIR | Sagittal T1 C+ | Axial T1 C+ | Axial T2 | Coronal T1 C+ |
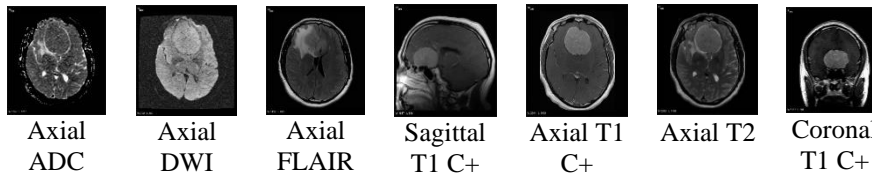
Figure 5. Dataset MRI image

The following is a sample of the experimental results from several images, as shown in Table 1.

Table 1. Sample of experiment results

Image Name	Original Image	Fast Local Laplacian	Color Map of K-Means FCM	Canny Edge Detection	Median Filter	Morphological Area Selection	Bounding Object
Mri1							
Mri2							
…	…	…	…	…	…	…	…
Mri29							
Mri30							

Image Name	Original Image	Fast Local Laplacian	Color Map of K-Means FCM	Canny Edge Detection	Median Filter	Morphological Area Selection	Bounding Object
Mri42							
...	...	...	...	...	...	...	...
Mri47							
Normal1							
...	...	...	...	...	...	...	...
Normal9							
...	...	...	...	...	...	...	...
Normal16							

The evaluation used is in the form of confusion matrix with the following conditions:

True Positive (TP): MRI image of a brain that has a tumor and a tumor is detected.

True Negative (TN): MRI images of the brain are normal and no tumor is detected.

False Positive (FP): MRI image of a brain that has a tumor but no tumor is detected.

False Negative (FN): MRI image of the brain is normal but a tumor is detected.

As a measure of the proposed method, the accuracy, precision, recall and error rate will be calculated. Here is the formula:

Accuracy = ((TP + TN)) / (Amount of data) × 100%; how accurate the system is in identifying the entire image correctly.

Error Rate = ((FP + FN)) / (Amount of data) × 100%; the percentage of errors that occur in the entire image.

Precision Positive = TP / ((TP + FP)) × 100%; accuracy of the MRI image of the brain that has a tumor and a tumor is detected.

Precision Negative = TN / ((TN + FN)) × 100%; accuracy of normal brain MRI images and no tumor detection.

Recall Positive = TP / ((TP + FN)) × 100%; the percentage value of categorical data from brain MRI images that have tumors and are correctly identified by the system.

Recall Negative = TN / ((TN + FP)) × 100%; percentage value of categorical data of normal brain MRI images that are correctly identified by the system.
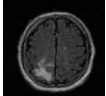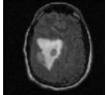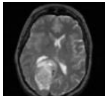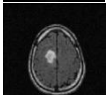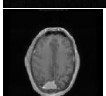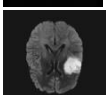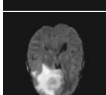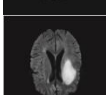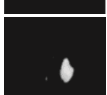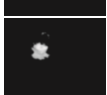
Table 2 is the calculation result of each component of the configuration matrix in the experimental data.

Table 2. Confussion matrix result

True Positive (TP)	52
True Negative (TN)	15
False Positive (FP)	5
False Negative (FN)	1
Accuracy	91.78%
Error Rate	8.22%
Precision Positive	91.23%
Precision Negative	93.75%
Recall Positive	98.11%
Recall Negative	75%

The comparison result with another research, as shown in Table 3.

Table 3. A Comparison Result with another Research

Original Image	Method [14]	Method [11]	K-Means FCM	K-Means FCM Without Canny
		-		
		-		
		-		
		-		
		-		
		-		
	-			
	-			
	-			
	-			

Experiments were also carried out on existing images in research by Saha [8] and Preetha [12]. The image is taken by cutting directly from the document using the "Snipping Tool" application, so the image obtained is of poor quality at a very small size. In addition, the type of MRI used is the Flair type, the detection of canny, which serves to thicken the line, is the main obstacle. The following is the result of the comparison. The accuracy obtained on the Flair type MRI is 100% without Canny or with Canny. It's just that, the detection results with Canny are not neat because the texture of the brain is considered a tumor. Conversely, if the MRI type T1 C + and T2, the results are better than if using Canny. If it is detected without using Canny, the accuracy will decrease.

In this study, more emphasis was placed on the Axial T1 C + and T2 MRI types. Experiments on Flair type MRI were used to prove that the K-Means FCM cluster can cluster properly. However, it is the steps after clustering that will determine the final result. With this, it is evident that the K-Means FCM cluster well. For the next experiment using only Axial T1 C + and T2 type MRIs.

CONCLUSION

Research on brain tumor detection on MRI images using the segmentation method has been carried out. The difficulty of determining an accurate method with a fast process is an obstacle to detecting brain tumors on MRI images. This is due to the large amount of noise present in the MRI image. In a study conducted on 73 MRI images of Axial T1 C + and T2 types where 57 MRI of the tumor brain and 16 MRI of the normal brain, plus 60 images of Flair type MRI. The K-Means method has a drawback, namely the clustering results are always changing. The FCM method has a drawback, namely the process is long because the calculation of each iteration is quite heavy. In combining these two methods, the results show that the shortcomings of the two methods can be minimized where K-Means acts as a preprocessing before the image is processed using FCM. The method of image enhancement (Image Enhancement) is also very influential in detecting objects. The use of the Fast Local Laplacian method is quite appropriate, because it can detail and adjust the contrast of the MRI image. The use of Canny's edge detection method also greatly influences the final result. Without an edge detection method, what will be detected is the area of the object and the intensity of its color. Meanwhile, if using the detection method, the detected edge is the circumference of the object. This is something that is quite interesting, because it has its respective advantages. If this study dosn't use the Canny method, the results will be more sensitive to color differences. If the research uses the Canny method, the result will be more sensitive to the lines that surround it. The use of the noise removal method using the median filter is chosen because it can replace pixels that are different from the others. This method is used to remove noise pixels around the line so that the line looks neater. The use of Morphological Area Selection is very important to select objects with a certain size. In this experiment, the most effective figure was 1/350 of the size of the MRI image. So that objects whose size is smaller than that size will be eliminated. In this research, the proposed method was found to get maximum results on the MRI dataset with the Flair type.

REFERENCES

[1]   S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain," *Egypt. J. Radiol. Nucl. Med.*, vol. 46, no. 2, pp. 475–479, 2015.

[2]   E. Küçükkülahli, P. Erdoğmuş, and K. Polat, "Brain MRI Segmentation based on Different Clustering Algorithms," *Int. J. Comput. Appl.*, vol. 155, no. 3, pp. 37–40, 2016.

[3]   U. Baid, S. Talbar, and S. Talbar, "Comparative Study of K-Means, Gaussian Mixture Models, Fuzzy-C-means algorithms for Brain Tumor Segmentation," *Advances in Intelligent Systems Research*, vol. 137, pp. 592-597, 2017.

[4]   C. C Benson, V. Deppa, V. L Lajish, and Kumar Rajamani, "Brain tumor segmentation from MR brain images using improved fuzzy c-means clustering and watershed algorithm," in *2016 Int. Conf. Adv. Comput. Commun. Inform (ICACCI)*, Jaipur, 2016.

[5]   P. Minal and S. Jankiballabh, "Brain Tumor Segmentation from MRI using Fractional Sobel Mask and Watershed Transform," *IEEE, Int. Conf. Inf. Commun. Instrum. Control (ICICIC)*, Indore, 2017.

[6]   Soumi Ghosh and Sanjay Kumar Dubey , "Comparative Analysis of K-Means and Fuzzy CMeans Algorithms," *(IJACSA) Int. J. Adv. Comput. Sci. Appl.* vol. 4, no. 4, pp. 35-39, 2013.

[7]   Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *J. Agric. Informatics*, vol. 6, no. 3, pp. 13–23, 2015.

[8]   C. Saha and M. F. Hossain, "MRI brain tumor images classification using K-means clustering, NSCT and SVM," *2017 4th IEEE Uttar Pradesh Sect. Int. Conf. Electr. Comput. Electron. UPCON 2017*, pp. 329–333, 2017.

[9]   C. Ji, K. Yin, T. Zhu, and C. Bian, "Canny Algorithm for High-speed Railway Apparatus Identification," *10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics (CISP-BMEI 2017)*, vol. 4, 2017.

[10]   D. E. P. Manullang, "Penyisipan Pesan Ke Dalam File Video Menerapkan Metode Chinese Remainder Theorem," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 108–117, 2019.

[11]   M. F. Dempsey, B. R. Condon, and D. M. Hadley, "Measurement of tumor 'size' in recurrent malignant glioma: 1D, 2D, or 3D?," *Am. J. Neuroradiol.*, vol. 26, no. 4, pp. 770–776, 2005.

[12]   R. Preetha and G. R. Suresh, "Performance analysis of fuzzy C means algorithm in automated detection of brain tumor," *Proc. - 2014 World Congr. Comput. Commun. Technol. WCCCT 2014*, no. 1, pp. 30–33, 2014.